

PICTURE PREDICTION METHOD AND RELATED APPARATUS

TECHNICAL FIELD

[0001] The present invention relates to the field of video coding/decoding, and in particular, to a picture prediction method and a related device.

5

BACKGROUND

[0002] With development of photoelectric acquisition technologies and increasing requirements for high-definition digital videos, an amount of video data is becoming large. Due to limited heterogeneous transmission bandwidths and diversified video applications, higher requirements are continuously imposed on video coding efficiency. Development of a high efficiency video coding (high efficiency video coding, HEVC for short) standard is initiated according to the requirements.

[0003] A basic principle of video compression coding is to use correlation between a space domain, a time domain, and a code word to remove redundancy as much as possible. Currently, a prevalent practice is to use a block-based hybrid video coding framework to implement video compression coding by performing steps of prediction (including intra-frame prediction and inter-frame prediction), transform, quantization, entropy coding, and the like. This coding framework shows high viability, and therefore, HEVC still uses this block-based hybrid video coding framework.

[0004] In various video coding/decoding solutions, motion estimation or motion compensation is a key technology that affects coding/decoding efficiency. In various conventional video coding/decoding solutions, it is assumed that motion of an object is always translational motion, and that motion of all parts of the entire object is the same. Basically, all conventional motion estimation or motion compensation algorithms are block motion compensation algorithms that are established based on a translational motion model (translational motion model). However, motion in the real world is diversified, and irregular motion such as scaling up/down, rotation, or parabolic motion is ubiquitous. Since the ninth decade of the last century, video coding experts have realized universality of irregular motion, and wished to introduce an irregular motion model (for example, an affine motion model) to improve video coding efficiency. However, computational complexity of conventional picture prediction performed based on the affine motion model is

implementations of the first aspect, in a twelfth possible implementation of the first aspect,

the affine motion model is in the following form:

$$\begin{cases} vx = \frac{vx_1 - vx_0}{w}x - \frac{vy_1 - vy_0}{w}y + vx_0 \\ vy = \frac{vy_1 - vy_0}{w}x + \frac{vx_1 - vx_0}{w}y + vy_0 \end{cases}, \text{ where}$$

the motion vectors of the two pixel samples are (vx_0, vy_0) and (vx_1, vy_1) respectively, vx

5 is a horizontal component of a motion vector of a pixel sample whose coordinates are (x, y) in the current picture block, vy is a vertical component of the motion vector of the pixel sample whose coordinates are (x, y) in the current picture block, and w is the length or width of the current picture block.

[0019] With reference to any one of the first aspect, or the first to the twelfth possible
10 implementations of the first aspect, in a thirteenth possible implementation of the first aspect,

the picture prediction method is applied to a video coding process, or the picture prediction method is applied to a video decoding process.

[0020] With reference to the thirteenth possible implementation of the first aspect, in a
fourteenth possible implementation of the first aspect, when the picture prediction method is
15 applied to the video decoding process, the determining, from N candidate merged motion information unit sets, the merged motion information unit set i including the two motion information units includes: determining, from the N candidate merged motion information unit sets, based on an identifier that is of the merged motion information unit set i and is obtained from a video bit stream, the merged motion information unit set i including the two motion information
20 units.

[0021] With reference to the thirteenth possible implementation of the first aspect or the
fourteenth possible implementation of the first aspect, in a fifteenth possible implementation of the
first aspect, when the picture prediction method is applied to the video decoding process, the
method further includes: decoding the video bit stream to obtain motion vector differences of the
25 two pixel samples, obtaining motion vector predictors of the two pixel samples by using motion vectors of pixel samples spatially adjacent or temporally adjacent to the two pixel samples, and obtaining the motion vectors of the two pixel samples separately based on the motion vector predictors of the two pixel samples and the motion vector differences of the two pixel samples.

[0022] With reference to the thirteenth possible implementation of the first aspect, in a sixteenth
possible implementation of the first aspect, when the picture prediction method is applied to the
30 video coding process, the method further includes: obtaining motion vector predictors of the two pixel samples by using motion vectors of pixel samples spatially adjacent or temporally adjacent to the two pixel samples, obtaining motion vector differences of the two pixel samples according to

fifth condition, where

the first condition includes that a motion mode of the current picture block indicated by a motion information unit in any one of the N candidate merged motion information unit sets is non-translational motion;

5 the second condition includes that prediction directions corresponding to two motion information units in any one of the N candidate merged motion information unit sets are the same;

the third condition includes that reference frame indexes corresponding to two motion information units in any one of the N candidate merged motion information unit sets are the same;

10 the fourth condition includes that an absolute value of a difference between horizontal components of motion vectors of two motion information units in any one of the N candidate merged motion information unit sets is less than or equal to a horizontal component threshold, or that an absolute value of a difference between a horizontal component of a motion vector of one motion information unit in any one of the N candidate merged motion information unit sets and a horizontal component of a motion vector of a pixel sample Z is less than or equal to a horizontal
15 component threshold, where the pixel sample Z in the current picture block is different from either of the two pixel samples; and

the fifth condition includes that an absolute value of a difference between vertical components of motion vectors of two motion information units in any one of the N candidate merged motion information unit sets is less than or equal to a vertical component threshold, or that
20 an absolute value of a difference between a vertical component of a motion vector of one motion information unit in any one of the N candidate merged motion information unit sets and a vertical component of a motion vector of a pixel sample Z is less than or equal to a vertical component threshold, where the pixel sample Z in the current picture block is different from either of the two pixel samples.

25 **[0027]** With reference to any one of the second aspect, or the first to the second possible implementations of the second aspect, in a third possible implementation of the second aspect, the two pixel samples include two pixel samples in an upper left pixel sample, an upper right pixel sample, a lower left pixel sample, and a central pixel sample a1 of the current picture block, where

the upper left pixel sample of the current picture block is an upper left vertex of the
30 current picture block, or a pixel block in the current picture block and including an upper left vertex of the current picture block; the lower left pixel sample of the current picture block is a lower left vertex of the current picture block, or a pixel block in the current picture block and including a lower left vertex of the current picture block; the upper right pixel sample of the current picture block is an upper right vertex of the current picture block, or a pixel block in the current picture

motion information unit set i.

[0033] With reference to any one of the second aspect, or the first to the eighth possible implementations of the second aspect, in a ninth possible implementation of the second aspect,

the predicting unit is specifically configured to: obtain a motion vector of each pixel in the current picture block through computation by using the affine motion model and the merged motion information unit set i, and determine a predicted pixel value of each pixel in the current picture block by using the motion vector that is of each pixel in the current picture block and is obtained through computation; or

the predicting unit is specifically configured to: obtain a motion vector of each pixel block in the current picture block through computation by using the affine motion model and the merged motion information unit set i, and determine a predicted pixel value of each pixel in each pixel block in the current picture block by using the motion vector that is of each pixel block in the current picture block and is obtained through computation.

[0034] With reference to any one of the second aspect, or the first to the ninth possible implementations of the second aspect, in a tenth possible implementation of the second aspect,

the predicting unit is specifically configured to obtain a motion vector of any pixel sample in the current picture block by using a ratio of a difference between horizontal components of motion vectors of the two pixel samples to a length or width of the current picture block and a ratio of a difference between vertical components of the motion vectors of the two pixel samples to the length or width of the current picture block, where the motion vectors of the two pixel samples are obtained based on motion vectors of the two motion information units in the merged motion information unit set i.

[0035] With reference to the tenth possible implementation of the second aspect, in an eleventh possible implementation of the second aspect, a horizontal coordinate coefficient of the horizontal components of the motion vectors of the two pixel samples is equal to a vertical coordinate coefficient of the vertical components of the motion vectors, and a vertical coordinate coefficient of the horizontal components of the motion vectors of the two pixel samples is opposite to a horizontal coordinate coefficient of the vertical components of the motion vectors.

[0036] With reference to any one of the second aspect, or the first to the eleventh possible implementations of the second aspect, in a twelfth possible implementation of the second aspect,

the affine motion model is in the following form:

$$\begin{cases} vx = \frac{vx_1 - vx_0}{w}x - \frac{vy_1 - vy_0}{w}y + vx_0 \\ vy = \frac{vy_1 - vy_0}{w}x + \frac{vx_1 - vx_0}{w}y + vy_0 \end{cases}, \text{ where}$$

the motion vectors of the two pixel samples are (vx_0, vy_0) and (vx_1, vy_1) respectively, vx

aspect of predicting the pixel value of the current picture block by using the affine motion model and the merged motion information unit set i, the processor is configured to: obtain a motion vector of each pixel in the current picture block through computation by using the affine motion model and the merged motion information unit set i, and determine a predicted pixel value of each pixel in the current picture block by using the motion vector that is of each pixel in the current picture block and is obtained through computation; or

in the aspect of predicting the pixel value of the current picture block by using the affine motion model and the merged motion information unit set i, the processor is configured to: obtain a motion vector of each pixel block in the current picture block through computation by using the affine motion model and the merged motion information unit set i, and determine a predicted pixel value of each pixel in each pixel block in the current picture block by using the motion vector that is of each pixel block in the current picture block and is obtained through computation.

[0052] With reference to any one of the third aspect, or the first to the ninth possible implementations of the third aspect, in a tenth possible implementation of the third aspect,

in the aspect of predicting the pixel value of the current picture block by using the affine motion model and the merged motion information unit set i, the processor is configured to obtain a motion vector of any pixel sample in the current picture block by using a ratio of a difference between horizontal components of motion vectors of the two pixel samples to a length or width of the current picture block and a ratio of a difference between vertical components of the motion vectors of the two pixel samples to the length or width of the current picture block, where the motion vectors of the two pixel samples are obtained based on motion vectors of the two motion information units in the merged motion information unit set i.

[0053] With reference to the tenth possible implementation of the third aspect, in an eleventh possible implementation of the third aspect,

a horizontal coordinate coefficient of the horizontal components of the motion vectors of the two pixel samples is equal to a vertical coordinate coefficient of the vertical components of the motion vectors, and a vertical coordinate coefficient of the horizontal components of the motion vectors of the two pixel samples is opposite to a horizontal coordinate coefficient of the vertical components of the motion vectors.

[0054] With reference to any one of the third aspect, or the first to the eleventh possible implementations of the third aspect, in a twelfth possible implementation of the third aspect,

the affine motion model is in the following form:

$$\begin{cases} vx = \frac{vx_1 - vx_0}{w}x - \frac{vy_1 - vy_0}{w}y + vx_0 \\ vy = \frac{vy_1 - vy_0}{w}x + \frac{vx_1 - vx_0}{w}y + vy_0 \end{cases}, \text{ where}$$

[0063] With reference to the fourth aspect, or the first or the second possible implementation of the fourth aspect, in a third possible implementation of the fourth aspect, the obtaining a motion vector of any pixel sample in the current picture block through computation by using an affine motion model and the motion vector 2-tuple includes: obtaining the values of the coefficients of the affine motion model by using a ratio of a difference between horizontal components of the motion vectors of the two pixel samples to a distance between the two pixel samples and a ratio of a difference between vertical components of the motion vectors of the two pixel samples to the distance between the two pixel samples; and

obtaining the motion vector of the any pixel sample in the current picture block by using the affine motion model and the values of the coefficients of the affine motion model.

[0064] With reference to the fourth aspect, or the first or the second possible implementation of the fourth aspect, in a fourth possible implementation of the fourth aspect, the obtaining a motion vector of any pixel sample in the current picture block through computation by using an affine motion model and the motion vector 2-tuple includes:

obtaining the values of the coefficients of the affine motion model by using a ratio of a weighted sum of components of the motion vectors of the two pixel samples to a distance between the two pixel samples or to a square of a distance between the two pixel samples; and

obtaining the motion vector of the any pixel sample in the current picture block by using the affine motion model and the values of the coefficients of the affine motion model.

[0065] With reference to any one of the fourth aspect, or the first to the third possible implementations of the fourth aspect, in a fifth possible implementation of the fourth aspect, when the two pixel samples include an upper left pixel sample of the current picture block and a pixel sample in a right region on a right side of the upper left pixel sample, the affine motion model is specifically:

$$\begin{cases} vx = \frac{vx_1 - vx_0}{w}x - \frac{vy_1 - vy_0}{w}y + vx_0 \\ vy = \frac{vy_1 - vy_0}{w}x + \frac{vx_1 - vx_0}{w}y + vy_0 \end{cases}, \text{ where}$$

(vx_0, vy_0) are a motion vector of the upper left pixel sample, (vx_1, vy_1) are a motion vector of the pixel sample in the right region, and w is a distance between the two pixel samples.

[0066] With reference to any one of the fourth aspect, or the first to the third possible implementations of the fourth aspect, in a sixth possible implementation of the fourth aspect, when the two pixel samples include an upper left pixel sample of the current picture block and a pixel sample in a lower region below the upper left pixel sample, the affine motion model is specifically:

$$\begin{cases} vx = \frac{vy_2 - vy_0}{h}x + \frac{vx_2 - vx_0}{h}y + vx_0 \\ vy = -\frac{vx_2 - vx_0}{h}x + \frac{vy_2 - vy_0}{h}y + vy_0 \end{cases}, \text{ where}$$

used in a conventional technology for screening out motion information units of multiple pixel samples only by performing a huge amount of computation in all possible candidate motion information unit sets corresponding to the multiple pixel samples is abandoned. This helps improve coding efficiency, also helps reduce computational complexity of picture prediction performed based on the affine motion model, further makes it possible to introduce the affine motion model into a video coding standard, and because the affine motion model is introduced, helps describe motion of an object more accurately, and therefore helps improve prediction accuracy. In addition, because a quantity of reference pixel samples may be two, this helps further reduce the computational complexity of picture prediction performed based on the affine motion model after the affine motion model is introduced, and also helps reduce affine parameter information or a quantity of motion vector differences or the like transferred by an encoder.

BRIEF DESCRIPTION OF DRAWINGS

[0093] To describe the technical solutions in the embodiments of the present invention more clearly, the following briefly describes the accompanying drawings required for describing the embodiments and the prior art. Apparently, the accompanying drawings in the following description show merely some embodiments of the present invention, and persons of ordinary skill in the art may still derive other drawings from these accompanying drawings without creative efforts.

[0094] FIG. 1-a and FIG. 1-b are schematic diagrams of several picture block partitioning modes according to an embodiment of the present invention;

[0095] FIG. 1-c is a schematic flowchart of a picture prediction method according to an embodiment of the present invention;

[0096] FIG. 1-d is a schematic diagram of a picture block according to an embodiment of the present invention;

[0097] FIG. 2-a is a schematic flowchart of another picture prediction method according to an embodiment of the present invention;

[0098] FIG. 2-b to FIG. 2-d are schematic diagrams of several methods for determining candidate motion information unit sets corresponding to pixel samples according to an embodiment of the present invention;

[0099] FIG. 2-e is a schematic diagram of vertex coordinates of a picture block x according to an embodiment of the present invention;

[0100] FIG. 2-f and FIG. 2-g are schematic diagrams of affine motion of a pixel according to an embodiment of the present invention;

[0101] FIG. 2-h and FIG. 2-i are schematic diagrams of rotational motion of a pixel according

without creative efforts shall fall within the protection scope of the present invention.

[0115] In the specification, claims, and accompanying drawings of the present invention, the terms "first", "second", "third", "fourth", and so on are intended to distinguish between different objects but do not indicate a particular order. In addition, the terms "including", "including", or any other variant thereof, are intended to cover a non-exclusive inclusion. For example, a process, a method, a system, a product, or a device that includes a series of steps or units is not limited to the listed steps or units, but optionally further includes an unlisted step or unit, or optionally further includes another inherent step or unit of the process, the method, the product, or the device.

[0116] The following first describes some concepts that may be used in the embodiments of the present invention.

[0117] In most coding frameworks, a video sequence includes a series of pictures (picture), the pictures are further divided into slices (slice), and the slices are further divided into blocks (block). Video coding is to perform coding processing from left to right and from top to bottom row by row starting from an upper left corner location of a picture by using a block as a unit. In some new video coding standards, the concept of a block is further extended. A macroblock (macroblock, MB for short) is defined in the H.264 standard, and the MB may be further divided into multiple prediction blocks (partition) that can be used for predictive coding. In the HEVC standard, basic concepts such as a coding unit (coding unit, CU for short), a prediction unit (prediction unit, PU for short), and a transform unit (transform unit, TU for short) are used, multiple units are classified according to functions, and a completely new tree-based structure is used for description. For example, the CU may be divided into smaller CUs according to a quadtree, and the smaller CUs may be further divided to form a quadtree structure. The PU and the TU also have similar tree structures. Regardless of whether a unit is a CU, a PU, or a TU, the unit belongs to the concept of a block block in essence. The CU is similar to a macroblock MB or a coding block, and is a basic unit for partitioning and encoding a coding picture. The PU may correspond to a prediction block, and is a basic unit for predictive coding. The CU is further divided into multiple PUs according to a partitioning mode. The TU may correspond to a transform block, and is a basic unit for transforming a prediction residual. In the high efficiency video coding (high efficiency video coding, HEVC for short) standard, they may be collectively referred to as coding tree blocks (coding tree block, CTB for short), or the like.

[0118] In the HEVC standard, a size of a coding unit may include four levels: 64x64, 32x32, 16x16, and 8x8. Coding units at each level may be divided into prediction units of different sizes according to intra-frame prediction and inter-frame prediction. For example, as shown in FIG. 1-a

and FIG. 1-b, FIG. 1-a illustrates by way of example a prediction unit partitioning mode corresponding to intra-frame prediction. FIG. 1-b illustrates by way of example several prediction unit partitioning modes corresponding to inter-frame prediction.

[0119] In a development and evolution process of a video coding technology, video coding experts figure out various methods to use temporal and spatial correlation between adjacent coding/decoding blocks to try to improve coding efficiency. In the H264 or advanced video coding (advanced video coding, AVC for short) standard, a skip mode (skip mode) and a direct mode (direct mode) become effective tools for improving coding efficiency. Blocks of the two coding modes used when a bit rate is low can occupy more than a half of an entire coding sequence. When the skip mode is used, a motion vector of a current picture block can be derived by using adjacent motion vectors only by adding a skip mode flag to a bit stream, and a value of a reference block is directly copied according to the motion vector as a reconstruction value of the current picture block. In addition, when the direct mode is used, an encoder may derive the motion vector of the current picture block by using the adjacent motion vectors, and directly copy the value of the reference block according to the motion vector as a predicted value of the current picture block, and perform predictive coding on the current picture block by using the predicted value in the encoder. In the current latest high efficiency video coding (high efficiency video coding, HEVC for short) standard, some new coding tools are introduced to further improve video coding efficiency. A merge coding (merge) mode and an advanced motion vector prediction (advanced motion vector prediction, AMVP for short) mode are two important inter-frame prediction tools. In merge coding (merge), motion information (which may include a motion vector (motion vector, MV for short), a prediction direction, a reference frame index, and the like) of coded blocks near a current coding block is used to construct a candidate motion information set; through comparison, candidate motion information with highest coding efficiency may be selected as motion information of the current coding block, a predicted value of the current coding block is found from the reference frame, and predictive coding is performed on the current coding block; and at a same time, an index value indicating from which adjacent coded block the motion information is selected may be written into a bit stream. When the advanced motion vector prediction mode is used, a motion vector of an adjacent coded block is used as a motion vector predictor of the current coding block. A motion vector with highest coding efficiency may be selected and used to predict a motion vector of the current coding block, and an index value indicating which adjacent motion vector is selected may be written a video bit stream.

[0120] The following continues to discuss the technical solutions of the embodiments of the present invention.

the pixel A1 (upper left central pixel), the pixel A2 (lower left central pixel), the pixel A3 (upper right central pixel), or the pixel A4 (lower right central pixel). Other cases may be deduced in the same way.

[0132] S102. Determine a merged motion information unit set *i* including two motion information units.

[0133] Each motion information unit in the merged motion information unit set *i* is selected from at least some of motion information units in the candidate motion information unit set corresponding to each of the two pixel samples. The motion information unit includes a motion vector whose prediction direction is forward and/or a motion vector whose prediction direction is backward.

[0134] For example, it is assumed that the two pixel samples include a pixel sample 001 and a pixel sample 002. A candidate motion information unit set corresponding to the pixel sample 001 is a candidate motion information unit set 011. A candidate motion information unit set corresponding to the pixel sample 002 is a candidate motion information unit set 022. The merged motion information unit set *i* includes a motion information unit C01 and a motion information unit C02. The motion information unit C01 may be selected from the candidate motion information unit set 011, the motion information unit C02 may be selected from the candidate motion information unit set 022, and so on.

[0135] It may be understood that, assuming that the merged motion information unit set *i* includes the motion information unit C01 and the motion information unit C02, where either of the motion information unit C01 and the motion information unit C02 may include a motion vector whose prediction direction is forward and/or a motion vector whose prediction direction is backward, the merged motion information unit set *i* may include two motion vectors (prediction directions corresponding to the two motion vectors may be forward or backward, or the two motion vectors may include one motion vector whose prediction direction is forward and one motion vector whose prediction direction is backward, or may include four motion vectors (the four motion vectors may include two motion vectors whose prediction directions are forward and two motion vectors whose prediction directions are backward), or may include three motion vectors (the three motion vectors may also include one motion vector whose prediction direction is forward and two motion vectors whose prediction directions are backward, or may include two motion vectors whose prediction directions are forward and one motion vector whose prediction direction is backward).

[0136] S103. Predict a pixel value of the current picture block by using an affine motion model and the merged motion information unit set *i*.

[0148] Optionally, in some possible implementations of the present invention, the determining, from N candidate merged motion information unit sets, the merged motion information unit set i including the two motion information units may include: determining, from the N candidate merged motion information unit sets, based on distortion or a rate distortion cost, the merged motion information unit set i including the two motion information units.

[0149] Optionally, a rate distortion cost corresponding to the merged motion information unit set i is lower than or equal to a rate distortion cost corresponding to any one of the N candidate merged motion information unit sets except the merged motion information unit set i.

[0150] Optionally, distortion corresponding to the merged motion information unit set i is less than or equal to distortion corresponding to any one of the N candidate merged motion information unit sets except the merged motion information unit set i.

[0151] A rate distortion cost corresponding to a candidate merged motion information unit set in the N candidate merged motion information unit sets (for example, the merged motion information unit set i in the N candidate merged motion information unit sets) may be, for example, a rate distortion cost corresponding to a predicted pixel value of a picture block (for example, the current picture block) obtained by predicting the pixel value of the picture block by using the candidate merged motion information unit set (for example, the merged motion information unit set i).

[0152] Distortion corresponding to a candidate merged motion information unit set in the N candidate merged motion information unit sets (for example, the merged motion information unit set i in the N candidate merged motion information unit sets) may be, for example, distortion between an original pixel value of a picture block (for example, the current picture block) and a predicted pixel value of the picture block obtained by predicting the pixel value of the picture block by using the candidate merged motion information unit set (for example, the merged motion information unit set i) (namely, the distortion between the original pixel value and the predicted pixel value of the picture block).

[0153] In some possible implementations of the present invention, distortion between an original pixel value of a picture block (for example, the current picture block) and a predicted pixel value of the picture block obtained by predicting the pixel value of the picture block by using the candidate merged motion information unit set (for example, the merged motion information unit set i) may be specifically, for example, a sum of quadratic differences (SSD, sum of quadratic differences) or a sum of absolute differences (SAD, sum of absolute differences) or a sum of differences between the original pixel value of the picture block (for example, the current picture block) and the predicted pixel value of the picture block obtained by predicting the pixel value of the picture block by using the candidate merged motion information unit set (for example, the

component of a motion vector of a pixel sample Z is less than or equal to a horizontal component threshold, where the pixel sample Z in the current picture block is different from either of the two pixel samples. The vertical component threshold may be, for example, equal to 1/3 of a height of the current picture block, 1/2 of a height of the current picture block, 2/3 of a height of the current picture block, 3/4 of a height of the current picture block, or another value.

[0164] Assuming that the two pixel samples are the upper left pixel sample and the upper right pixel sample of the current picture block, the pixel sample Z may be the lower left pixel sample, or the central pixel sample, or any other pixel sample of the current picture block. Other cases may be deduced in the same way.

[0165] Optionally, in some possible implementations of the present invention, a candidate motion information unit set corresponding to the upper left pixel sample of the current picture block includes motion information units of x1 pixel samples, where the x1 pixel samples include at least one pixel sample spatially adjacent to the upper left pixel sample of the current picture block and/or at least one pixel sample temporally adjacent to the upper left pixel sample of the current picture block, and x1 is a positive integer. For example, the x1 pixel samples include only at least one pixel sample spatially adjacent to the upper left pixel sample of the current picture block and/or at least one pixel sample temporally adjacent to the upper left pixel sample of the current picture block.

[0166] For example, x1 may be, for example, equal to 1, 2, 3, 4, 5, 6, or another value.

[0167] For example, the x1 pixel samples include at least one of a pixel sample that has a same location as the upper left pixel sample of the current picture block, a pixel sample spatially adjacent to a left edge of the current picture block, a pixel sample spatially adjacent to an upper left of the current picture block, or a pixel sample spatially adjacent to an upper edge of the current picture block, in a video frame temporally adjacent to a video frame to which the current picture block belongs.

[0168] Optionally, in some possible implementations of the present invention, a candidate motion information unit set corresponding to the upper right pixel sample of the current picture block includes motion information units of x2 pixel samples, where the x2 pixel samples include at least one pixel sample spatially adjacent to the upper right pixel sample of the current picture block and/or at least one pixel sample temporally adjacent to the upper right pixel sample of the current picture block, and x2 is a positive integer.

[0169] For example, x2 may be, for example, equal to 1, 2, 3, 4, 5, 6, or another value.

[0170] For example, the x2 pixel samples include at least one of a pixel sample that has a same location as the upper right pixel sample of the current picture block, a pixel sample spatially adjacent to a right edge of the current picture block, a pixel sample spatially adjacent to an upper right of the current picture block, or a pixel sample spatially adjacent to the upper edge of the

frame, the zero motion information unit added to the candidate motion information unit set S3 includes a zero motion vector whose prediction direction is forward but may not include a zero motion vector whose prediction direction is backward. If the video frame to which the current picture block belongs is a backward prediction frame, the zero motion information unit added to the candidate motion information unit set S3 includes a zero motion vector whose prediction direction is backward but may not include a zero motion vector whose prediction direction is forward. In addition, if the video frame to which the current picture block belongs is a bidirectional prediction frame, the zero motion information unit added to the candidate motion information unit set S3 includes a zero motion vector whose prediction direction is forward and a zero motion vector whose prediction direction is backward. Reference frame indexes corresponding to motion vectors in different zero motion information units added to the candidate motion information unit set S3 may be different, and the corresponding reference frame indexes may be, for example, 0, 1, 2, 3, or another value.

[0202] That two motion information units are different may mean that motion vectors included in the two motion information units are different, or that prediction directions corresponding to motion vectors included in the two motion information units are different, or that reference frame indexes corresponding to motion vectors included in the two motion information units are different. That two motion information units are the same may mean that motion vectors included in the two motion information units are the same, and that prediction directions corresponding to the motion vectors included in the two motion information units are the same, and that reference frame indexes corresponding to the motion vectors included in the two motion information units are the same.

[0203] It may be understood that, for a scenario in which more pixel samples exist, a candidate motion information unit set corresponding to a pixel sample may be obtained in a similar manner.

[0204] For example, as shown in FIG. 2-d, the two pixel samples may include two pixel samples in an upper left pixel sample, an upper right pixel sample, a lower left pixel sample, and a central pixel sample a1 of the current picture block. The upper left pixel sample of the current picture block is an upper left vertex of the current picture block, or a pixel block in the current picture block and including an upper left vertex of the current picture block; the lower left pixel sample of the current picture block is a lower left vertex of the current picture block, or a pixel block in the current picture block and including a lower left vertex of the current picture block; the upper right pixel sample of the current picture block is an upper right vertex of the current picture block, or a pixel block in the current picture block and including an upper right vertex of the current picture block; and the central pixel sample a1 of the current picture block is a central pixel of the current picture block, or a pixel block in the current picture block and

information unit set i is scaled down to a forward reference frame of the current picture block and that the motion vector whose prediction direction is backward in the merged motion information unit set i is scaled down to a backward reference frame of the current picture block, and predicting the pixel value of the current picture block by using the affine motion model and a scaled merged motion information unit set i.

[0212] S204. The video coding apparatus determines, from the N candidate merged motion information unit sets, a merged motion information unit set i including two motion information units.

[0213] Optionally, in some possible implementations of the present invention, the video coding apparatus may further write an identifier of the merged motion information unit set i into a video bit stream. Correspondingly, a video decoding apparatus determines, from the N candidate merged motion information unit sets, based on the identifier that is of the merged motion information unit set i and is obtained from the video bit stream, the merged motion information unit set i including the two motion information units.

[0214] Optionally, in some possible implementations of the present invention, that the video coding apparatus determines, from the N candidate merged motion information unit sets, a merged motion information unit set i including two motion information units may include: determining, from the N candidate merged motion information unit sets, based on distortion or a rate distortion cost, the merged motion information unit set i including two motion information units.

[0215] Optionally, a rate distortion cost corresponding to the merged motion information unit set i is lower than or equal to a rate distortion cost corresponding to any one of the N candidate merged motion information unit sets except the merged motion information unit set i.

[0216] Optionally, distortion corresponding to the merged motion information unit set i is less than or equal to distortion corresponding to any one of the N candidate merged motion information unit sets except the merged motion information unit set i.

[0217] A rate distortion cost corresponding to a candidate merged motion information unit set in the N candidate merged motion information unit sets (for example, the merged motion information unit set i in the N candidate merged motion information unit sets) may be, for example, a rate distortion cost corresponding to a predicted pixel value of a picture block (for example, the current picture block) obtained by predicting the pixel value of the picture block by using the candidate merged motion information unit set (for example, the merged motion information unit set i).

[0218] Distortion corresponding to a candidate merged motion information unit set in the N candidate merged motion information unit sets (for example, the merged motion information unit set i in the N candidate merged motion information unit sets) may be, for example, distortion

and therefore helps improve prediction accuracy. Because a quantity of reference pixel samples may be two, this helps further reduce the computational complexity of picture prediction performed based on the affine motion model after the affine motion model is introduced, and also helps reduce affine parameter information or a quantity of motion vector differences or the like transferred by an encoder.

[0232] The following illustrates a process of deriving the affine motion model shown in the formula 1. For example, a rotational motion model may be used to derive the affine motion model.

[0233] For example, rotational motion is shown by way of example in FIG. 2-h or FIG. 2-i.

[0234] The rotational motion model is shown in a formula 2, where (x', y') are coordinates corresponding to a pixel whose coordinates are (x, y) , in a reference frame, θ is an angle of rotation, and (a_0, a_1) are translational components. If a transform coefficient is known, a motion vector (vx, vy) of a pixel (x, y) may be obtained:

$$\begin{cases} x' = \cos \theta \cdot x + \sin \theta \cdot y + a_0 \\ y' = -\sin \theta \cdot x + \cos \theta \cdot y + a_1 \\ vx = x - x' \\ vy = y - y' \end{cases} \quad (\text{formula 2})$$

[0235] A used rotational matrix is $\begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix}$.

[0236] Based on rotation, if zooming is performed by using a coefficient ρ , the following simplified affine motion model is obtained, so as to avoid triangular computation in rotational motion:

$$\begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} \rho & 0 \\ 0 & \rho \end{bmatrix} = \begin{bmatrix} \rho \cos \theta & \rho \sin \theta \\ -\rho \sin \theta & \rho \cos \theta \end{bmatrix} = \begin{bmatrix} a_2 & a_3 \\ -a_3 & a_2 \end{bmatrix}$$

[0237] This helps reduce computational complexity, and may simplify a process of computing the motion vector of each pixel. In addition, like a general affine motion model, this model may be applied to complex motion scenarios such as rotation and scaling. The simplified affine motion model may be described in a formula 3. In comparison with the general affine motion model, only four parameters may be required for expressing the simplified affine motion model:

$$\begin{cases} x' = a_2 \cdot x + a_3 \cdot y + a_0 \\ y' = -a_3 \cdot x + a_2 \cdot y + a_1 \\ vx = x - x' \\ vy = y - y' \end{cases} \quad (\text{formula 3})$$

[0238] For a picture block (for example, CUR) whose size is $w \times h$, a right edge and a lower edge of the picture block are separately extended by one row to obtain vertices whose coordinates are $(0,0)$ and $(w,0)$, and motion vectors (vx_0,vy_0) and (vx_1,vy_1) of vertices whose

5 coordinates are $(0,0)$ and $(w,0)$ are obtained. The two vertices are used as pixel samples (certainly, another pixel may also be used as a reference pixel sample, for example, a central pixel sample), the coordinates and motion vectors of the pixel samples are substituted into the formula 3, and the formula 1 may be derived:

$$\begin{cases} vx = \frac{vx_1 - vx_0}{w} x - \frac{vy_1 - vy_0}{w} y + vx_0 \\ vy = \frac{vy_1 - vy_0}{w} x + \frac{vx_1 - vx_0}{w} y + vy_0 \end{cases} \quad (\text{formula 1})$$

$$\begin{cases} \frac{vx_2 - vx_0}{h} = -\frac{vy_1 - vy_0}{w} \\ \frac{vy_2 - vy_0}{h} = \frac{vx_1 - vx_0}{w} \end{cases} \quad \text{where}$$

[0239] The motion vectors of the two pixel samples are (vx_0,vy_0) and (vx_1,vy_1) respectively, vx is a horizontal component of a motion vector of a pixel sample whose coordinates are (x,y) in the current picture block, vy is a vertical component of the motion vector of the pixel sample whose coordinates are (x,y) in the current picture block, and w is a length or width of the current picture block.

15 **[0240]** It may be understood that, the foregoing derivation process shows that the formula 1 has relatively high availability. During practice, it is found that, because a quantity of reference pixel samples may be two, this helps further reduce the computational complexity of picture prediction performed based on the affine motion model after the affine motion model is introduced, and reduce affine parameter information or a quantity of motion vector differences transferred during coding.

20 **[0241]** Referring to FIG. 3, FIG. 3 is a schematic flowchart of another picture prediction method according to another embodiment of the present invention. This embodiment is described mainly by using a picture prediction method implemented in a video decoding apparatus as an example. As shown by way of example in FIG. 3, another picture prediction method provided by another embodiment of the present invention may include the following steps.

is backward but may not include a zero motion vector whose prediction direction is forward. In addition, if the video frame to which the current picture block belongs is a bidirectional prediction frame, the zero motion information unit added to the candidate motion information unit set S3 includes a zero motion vector whose prediction direction is forward and a zero motion vector whose prediction direction is backward. Reference frame indexes corresponding to motion vectors in different zero motion information units added to the candidate motion information unit set S3 may be different, and the corresponding reference frame indexes may be, for example, 0, 1, 2, 3, or another value.

[0255] That two motion information units are different may mean that motion vectors included in the two motion information units are different, or that prediction directions corresponding to motion vectors included in the two motion information units are different, or that reference frame indexes corresponding to motion vectors included in the two motion information units are different. That two motion information units are the same may mean that motion vectors included in the two motion information units are the same, and that prediction directions corresponding to the motion vectors included in the two motion information units are the same, and that reference frame indexes corresponding to the motion vectors included in the two motion information units are the same.

[0256] It may be understood that, for a scenario in which more pixel samples exist, a candidate motion information unit set corresponding to a pixel sample may be obtained in a similar manner.

[0257] For example, as shown in FIG. 2-d, the two pixel samples may include two pixel samples in an upper left pixel sample, an upper right pixel sample, a lower left pixel sample, and a central pixel sample a1 of the current picture block. The upper left pixel sample of the current picture block is an upper left vertex of the current picture block, or a pixel block in the current picture block and including an upper left vertex of the current picture block; the lower left pixel sample of the current picture block is a lower left vertex of the current picture block, or a pixel block in the current picture block and including a lower left vertex of the current picture block; the upper right pixel sample of the current picture block is an upper right vertex of the current picture block, or a pixel block in the current picture block and including an upper right vertex of the current picture block; and the central pixel sample a1 of the current picture block is a central pixel of the current picture block, or a pixel block in the current picture block and including a central pixel of the current picture block.

[0258] S303. The video decoding apparatus determines N candidate merged motion information unit sets based on the candidate motion information unit set corresponding to each of the two pixel samples. Each motion information unit included in each of the N candidate merged motion information unit sets is selected from at least some of constraint-compliant motion information units

using an affine motion model and the merged motion information unit set i.

[0278] Optionally, in some possible implementations of the present invention, the second determining unit 420 may be specifically configured to determine, from N candidate merged motion information unit sets, the merged motion information unit set i including the two motion information units, where each motion information unit included in each of the N candidate merged motion information unit sets is selected from at least some of constraint-compliant motion information units in the candidate motion information unit set corresponding to each of the two pixel samples, N is a positive integer, the N candidate merged motion information unit sets are different from each other, and each of the N candidate merged motion information unit sets includes two motion information units.

[0279] Optionally, in some possible implementations of the present invention, the N candidate merged motion information unit sets meet at least one of a first condition, a second condition, a third condition, a fourth condition, or a fifth condition, where

the first condition includes that a motion mode of the current picture block indicated by a motion information unit in any one of the N candidate merged motion information unit sets is non-translational motion;

the second condition includes that prediction directions corresponding to two motion information units in any one of the N candidate merged motion information unit sets are the same;

the third condition includes that reference frame indexes corresponding to two motion information units in any one of the N candidate merged motion information unit sets are the same;

the fourth condition includes that an absolute value of a difference between horizontal components of motion vectors of two motion information units in any one of the N candidate merged motion information unit sets is less than or equal to a horizontal component threshold, or that an absolute value of a difference between a horizontal component of a motion vector of one motion information unit in any one of the N candidate merged motion information unit sets and a horizontal component of a motion vector of a pixel sample Z is less than or equal to a horizontal component threshold, where the pixel sample Z in the current picture block is different from either of the two pixel samples; and

the fifth condition includes that an absolute value of a difference between vertical components of motion vectors of two motion information units in any one of the N candidate merged motion information unit sets is less than or equal to a vertical component threshold, or that an absolute value of a difference between a vertical component of a motion vector of one motion information unit in any one of the N candidate merged motion information unit sets and a vertical component of a motion vector of a pixel sample Z is less than or equal to a vertical component

threshold, where the pixel sample Z in the current picture block is different from either of the two pixel samples.

[0280] Optionally, in some possible implementations of the present invention, the two pixel samples include two pixel samples in an upper left pixel sample, an upper right pixel sample, a lower left pixel sample, and a central pixel sample a1 of the current picture block, where

the upper left pixel sample of the current picture block is an upper left vertex of the current picture block, or a pixel block in the current picture block and including an upper left vertex of the current picture block; the lower left pixel sample of the current picture block is a lower left vertex of the current picture block, or a pixel block in the current picture block and including a lower left vertex of the current picture block; the upper right pixel sample of the current picture block is an upper right vertex of the current picture block, or a pixel block in the current picture block and including an upper right vertex of the current picture block; and the central pixel sample a1 of the current picture block is a central pixel of the current picture block, or a pixel block in the current picture block and including a central pixel of the current picture block.

[0281] Optionally, in some possible implementations of the present invention, a candidate motion information unit set corresponding to the upper left pixel sample of the current picture block includes motion information units of x1 pixel samples, where the x1 pixel samples include at least one pixel sample spatially adjacent to the upper left pixel sample of the current picture block and/or at least one pixel sample temporally adjacent to the upper left pixel sample of the current picture block, and x1 is a positive integer, where

the x1 pixel samples include at least one of a pixel sample that has a same location as the upper left pixel sample of the current picture block, a pixel sample spatially adjacent to a left edge of the current picture block, a pixel sample spatially adjacent to an upper left of the current picture block, or a pixel sample spatially adjacent to an upper edge of the current picture block, in a video frame temporally adjacent to a video frame to which the current picture block belongs.

[0282] Optionally, in some possible implementations of the present invention, a candidate motion information unit set corresponding to the upper right pixel sample of the current picture block includes motion information units of x2 pixel samples, where the x2 pixel samples include at least one pixel sample spatially adjacent to the upper right pixel sample of the current picture block and/or at least one pixel sample temporally adjacent to the upper right pixel sample of the current picture block, and x2 is a positive integer, where

the x2 pixel samples include at least one of a pixel sample that has a same location as the upper right pixel sample of the current picture block, a pixel sample spatially adjacent to a right edge of the current picture block, a pixel sample spatially adjacent to an upper right of the current

index of the current picture block, perform scaling processing on the merged motion information unit set i, so that the motion vector whose prediction direction is forward in the merged motion information unit set i is scaled down to a forward reference frame of the current picture block and that the motion vector whose prediction direction is backward in the merged motion information unit set i is scaled down to a backward reference frame of the current picture block, and predict the pixel value of the current picture block by using the affine motion model and a scaled merged motion information unit set i.

[0286] Optionally, in some possible implementations of the present invention, the predicting unit 430 is specifically configured to: obtain a motion vector of each pixel in the current picture block through computation by using the affine motion model and the merged motion information unit set i, and determine a predicted pixel value of each pixel in the current picture block by using the motion vector that is of each pixel in the current picture block and is obtained through computation; or

the predicting unit 430 is specifically configured to: obtain a motion vector of each pixel block in the current picture block through computation by using the affine motion model and the merged motion information unit set i, and determine a predicted pixel value of each pixel in each pixel block in the current picture block by using the motion vector that is of each pixel block in the current picture block and is obtained through computation.

[0287] Optionally, in some possible implementations of the present invention, the predicting unit 430 may be specifically configured to obtain a motion vector of any pixel sample in the current picture block by using a ratio of a difference between horizontal components of motion vectors of the two pixel samples to a length or width of the current picture block and a ratio of a difference between vertical components of the motion vectors of the two pixel samples to the length or width of the current picture block, where the motion vectors of the two pixel samples are obtained based on motion vectors of the two motion information units in the merged motion information unit set i.

[0288] Optionally, in some possible implementations of the present invention, a horizontal coordinate coefficient of the horizontal components of the motion vectors of the two pixel samples is equal to a vertical coordinate coefficient of the vertical components of the motion vectors, and a vertical coordinate coefficient of the horizontal components of the motion vectors of the two pixel samples is opposite to a horizontal coordinate coefficient of the vertical components of the motion vectors.

[0289] Optionally, in some possible implementations of the present invention, the affine motion model may be in the following form:

encoding unit, configured to write an identifier of the merged motion information unit set i into the video bit stream.

[0295] It may be understood that, functions of functional units of the picture prediction apparatus 400 in this embodiment may be specifically implemented according to the method in the foregoing method embodiment. For a specific implementation process thereof, refer to the related description in the foregoing method embodiment. Details are not described again herein. The picture prediction apparatus 400 may be any apparatus that needs to output and play a video, for example, a device such as a notebook computer, a tablet computer, a personal computer, or a mobile phone.

[0296] It can be learned that, in the technical solution provided by this embodiment, the picture prediction apparatus 400 predicts a pixel value of a current picture block by using an affine motion model and a merged motion information unit set i, where each motion information unit in the merged motion information unit set i is selected from at least some of motion information units in a candidate motion information unit set corresponding to each of two pixel samples. Because a selection range of the merged motion information unit set i becomes relatively small, a mechanism used in a conventional technology for screening out motion information units of multiple pixel samples only by performing a huge amount of computation in all possible candidate motion information unit sets of the multiple pixel samples is abandoned. This helps improve coding efficiency, also helps reduce computational complexity of picture prediction performed based on the affine motion model, further makes it possible to introduce the affine motion model into a video coding standard, and because the affine motion model is introduced, helps describe motion of an object more accurately, and therefore helps improve prediction accuracy. In addition, because a quantity of reference pixel samples may be two, this helps further reduce the computational complexity of picture prediction performed based on the affine motion model after the affine motion model is introduced, and also helps reduce affine parameter information or a quantity of motion vector differences or the like transferred by an encoder.

[0297] Referring to FIG. 5, FIG. 5 is a schematic diagram of a picture prediction apparatus 500 according to an embodiment of the present invention. The picture prediction apparatus 500 may include at least one bus 501, at least one processor 502 connected to the bus 501, and at least one memory 503 connected to the bus 501.

[0298] The processor 502 invokes, by using the bus 501, code or an instruction stored in the memory 503, so that the processor 502 is configured to: determine two pixel samples in a current picture block, and determine a candidate motion information unit set corresponding to each of the two pixel samples, where the candidate motion information unit set corresponding to each pixel

the fifth condition includes that an absolute value of a difference between vertical components of motion vectors of two motion information units in any one of the N candidate merged motion information unit sets is less than or equal to a vertical component threshold, or that an absolute value of a difference between a vertical component of a motion vector of one motion information unit in any one of the N candidate merged motion information unit sets and a vertical component of a motion vector of a pixel sample Z is less than or equal to a vertical component threshold, where the pixel sample Z in the current picture block is different from either of the two pixel samples.

[0301] Optionally, in some possible implementations of the present invention, the two pixel samples include two pixel samples in an upper left pixel sample, an upper right pixel sample, a lower left pixel sample, and a central pixel sample a1 of the current picture block, where

the upper left pixel sample of the current picture block is an upper left vertex of the current picture block, or a pixel block in the current picture block and including an upper left vertex of the current picture block; the lower left pixel sample of the current picture block is a lower left vertex of the current picture block, or a pixel block in the current picture block and including a lower left vertex of the current picture block; the upper right pixel sample of the current picture block is an upper right vertex of the current picture block, or a pixel block in the current picture block and including an upper right vertex of the current picture block; and the central pixel sample a1 of the current picture block is a central pixel of the current picture block, or a pixel block in the current picture block and including a central pixel of the current picture block.

[0302] Optionally, in some possible implementations of the present invention, a candidate motion information unit set corresponding to the upper left pixel sample of the current picture block includes motion information units of x1 pixel samples, where the x1 pixel samples include at least one pixel sample spatially adjacent to the upper left pixel sample of the current picture block and/or at least one pixel sample temporally adjacent to the upper left pixel sample of the current picture block, and x1 is a positive integer, where

the x1 pixel samples include at least one of a pixel sample that has a same location as the upper left pixel sample of the current picture block, a pixel sample spatially adjacent to a left edge of the current picture block, a pixel sample spatially adjacent to an upper left of the current picture block, or a pixel sample spatially adjacent to an upper edge of the current picture block, in a video frame temporally adjacent to a video frame to which the current picture block belongs.

[0303] Optionally, in some possible implementations of the present invention, a candidate motion information unit set corresponding to the upper right pixel sample of the current picture block includes motion information units of x2 pixel samples, where the x2 pixel samples include at

block and a ratio of a difference between vertical components of the motion vectors of the two pixel samples to the length or width of the current picture block, where the motion vectors of the two pixel samples are obtained based on motion vectors of the two motion information units in the merged motion information unit set i.

[0309] Optionally, in some possible implementations of the present invention, a horizontal coordinate coefficient of the horizontal components of the motion vectors of the two pixel samples is equal to a vertical coordinate coefficient of the vertical components of the motion vectors, and a vertical coordinate coefficient of the horizontal components of the motion vectors of the two pixel samples is opposite to a horizontal coordinate coefficient of the vertical components of the motion vectors.

[0310] Optionally, in some possible implementations of the present invention, the affine motion model may be in the following form:

$$\begin{cases} vx = \frac{vx_1 - vx_0}{w} x - \frac{vy_1 - vy_0}{w} y + vx_0 \\ vy = \frac{vy_1 - vy_0}{w} x + \frac{vx_1 - vx_0}{w} y + vy_0 \end{cases}, \text{ where}$$

the motion vectors of the two pixel samples are (vx_0, vy_0) and (vx_1, vy_1) respectively, vx is a horizontal component of a motion vector of a pixel sample whose coordinates are (x, y) in the current picture block, vy is a vertical component of the motion vector of the pixel sample whose coordinates are (x, y) in the current picture block, and w is the length or width of the current picture block.

[0311] Optionally, in some possible implementations of the present invention, the picture prediction apparatus is applied to a video coding apparatus, or the picture prediction apparatus is applied to a video decoding apparatus.

[0312] Optionally, in some possible implementations of the present invention, when the picture prediction apparatus is applied to the video decoding apparatus, in the aspect of determining the merged motion information unit set i including the two motion information units, the processor 502 is configured to determine, from the N candidate merged motion information unit sets, based on an identifier that is of the merged motion information unit set i and is obtained from a video bit stream, the merged motion information unit set i including the two motion information units.

[0313] Optionally, in some possible implementations of the present invention, when the picture prediction apparatus is applied to the video decoding apparatus, the processor 502 is further configured to: decode the video bit stream to obtain motion vector differences of the two pixel samples, obtain motion vector predictors of the two pixel samples by using motion vectors of pixel samples spatially adjacent or temporally adjacent to the two pixel samples, and obtain the motion

vectors of the two pixel samples separately based on the motion vector predictors of the two pixel samples and the motion vector differences of the two pixel samples.

[0314] Optionally, in some possible implementations of the present invention, when the picture prediction apparatus is applied to the video coding apparatus, the processor 502 is further configured to: obtain motion vector predictors of the two pixel samples by using motion vectors of pixel samples spatially adjacent or temporally adjacent to the two pixel samples, obtain motion vector differences of the two pixel samples according to the motion vector predictors of the two pixel samples, and write the motion vector differences of the two pixel samples into a video bit stream.

[0315] Optionally, in some possible implementations of the present invention, when the picture prediction apparatus is applied to the video coding apparatus, the processor 502 is further configured to write an identifier of the merged motion information unit set *i* into the video bit stream.

[0316] It may be understood that, functions of functional units of the picture prediction apparatus 500 in this embodiment may be specifically implemented according to the method in the foregoing method embodiment. For a specific implementation process thereof, refer to the related description in the foregoing method embodiment. Details are not described again herein. The picture prediction apparatus 500 may be any apparatus that needs to output and play a video, for example, a device such as a notebook computer, a tablet computer, a personal computer, or a mobile phone.

[0317] It can be learned that, in the technical solution provided by this embodiment, the picture prediction apparatus 500 predicts a pixel value of a current picture block by using an affine motion model and a merged motion information unit set *i*, where each motion information unit in the merged motion information unit set *i* is selected from at least some of motion information units in a candidate motion information unit set corresponding to each of two pixel samples. Because a selection range of the merged motion information unit set *i* becomes relatively small, a mechanism used in a conventional technology for screening out motion information units of multiple pixel samples only by performing a huge amount of computation in all possible candidate motion information unit sets of the multiple pixel samples is abandoned. This helps improve coding efficiency, also helps reduce computational complexity of picture prediction performed based on the affine motion model, further makes it possible to introduce the affine motion model into a video coding standard, and because the affine motion model is introduced, helps describe motion of an object more accurately, and therefore helps improve prediction accuracy. In addition, because a quantity of reference pixel samples may be two, this helps further reduce the computational

[0325] The two pixel samples may include two pixel samples in an upper left pixel sample, a pixel sample in a right region, a pixel sample in a lower region, and a pixel sample in a lower right region of the current picture block.

[0326] The upper left pixel sample of the current picture block may be an upper left vertex of the current picture block, or a pixel block in the current picture block and including an upper left vertex of the current picture block. In video coding and decoding, coordinate values of the upper left pixel sample may be (0, 0) by default.

[0327] The pixel sample in the lower region of the current picture block may be a pixel or a pixel block below the upper left pixel sample of the current picture block. A vertical coordinate of the pixel sample in the lower region is greater than a vertical coordinate of the upper left pixel sample. The pixel sample in the lower region may include the lower left pixel sample in the foregoing embodiment. A horizontal coordinate of the pixel sample in the lower region may be the same as a horizontal coordinate of the upper left pixel sample, or a difference between a horizontal coordinate of the pixel sample in the lower region and a horizontal coordinate of the upper left pixel sample may be n pixels in height, where n is a positive integer less than 3.

[0328] The pixel sample in the right region of the current picture block may be a pixel or a pixel block on a right side of the upper left pixel sample of the current picture block. A horizontal coordinate of the pixel sample in the right region is greater than the horizontal coordinate of the upper left pixel sample. The pixel sample in the right region may include the upper right pixel sample in the foregoing embodiment. A vertical coordinate of the pixel sample in the right region may be the same as the vertical coordinate of the upper left pixel sample, or a difference between a vertical coordinate of the pixel sample in the right region and the vertical coordinate of the upper left pixel sample may be n pixels in width, where n is a positive integer less than 3.

[0329] The pixel sample in the lower right region of the current picture block may be a pixel or a pixel block on a lower right side of the upper left pixel sample of the current picture block. A vertical coordinate of the pixel sample in the lower right region is greater than the vertical coordinate of the upper left pixel sample, and a horizontal coordinate of the pixel sample in the lower right region is greater than the horizontal coordinate of the upper left pixel sample. The pixel sample in the lower right region may include the central pixel sample a_1 in the foregoing embodiment, and may further include a lower right pixel sample. The lower right pixel sample of the current picture block may be a lower right vertex of the current picture block, or a pixel block in the current picture block and including a lower right vertex of the current picture block.

predicted pixel value of each pixel in the current picture block by using the motion vector that is of each pixel in the current picture block and is obtained through computation. The motion vector of the any pixel sample in the current picture block may also be a motion vector of any pixel block in the current picture block, and the process may be: determining a predicted pixel value of each pixel
5 in each pixel block in the current picture block by using the motion vector that is of each pixel block in the current picture block and is obtained through computation.

[0348] During a test, it is found that performing motion vector computation by using a pixel block in the current picture block as a granularity helps reduce computational complexity greatly if the motion vector of each pixel block in the current picture block is obtained through computation
10 by using the affine motion model and the merged motion information unit set i and then the predicted pixel value of each pixel in each pixel block in the current picture block is determined by using the motion vector that is of each pixel block in the current picture block and is obtained through computation.

[0349] Optionally, in some possible implementations of the present invention, when the picture
15 processing method is applied to a video coding process, the method may further include: performing motion compensation predictive coding on the any pixel sample in the current picture block by using the motion vector that is of the any pixel sample in the current picture block and is obtained through computation.

[0350] Specifically, the process may be: determining the predicted pixel value of the pixel of
20 the any pixel sample in the current picture block by using the motion vector that is of the any pixel sample in the current picture block and is obtained through computation; and performing motion compensation prediction on the any pixel sample by using the predicted pixel value of the pixel of the any pixel sample, so as to obtain a pixel reconstruction value of the any pixel sample.

[0351] Alternatively, the process may be: determining the predicted pixel value of the pixel of
25 the any pixel sample in the current picture block by using the motion vector that is of the any pixel sample in the current picture block and is obtained through computation; and performing motion compensation prediction on the any pixel sample by using the predicted pixel value of the pixel of the any pixel sample, obtaining a prediction residual of the any pixel sample by using the pixel value of the pixel of the any pixel sample obtained by performing motion compensation prediction
30 and an actual pixel value of the pixel of the any pixel sample, and encoding the prediction residual of the any pixel sample into a bit stream.

[0352] Alternatively, after the prediction residual of the any pixel sample is obtained, a similar method is used to obtain prediction residuals of other pixel samples required by a prediction residual of the current picture block, so that the prediction residual of the current picture block is

obtained. Then the prediction residual of the current picture block is encoded into the bit stream. The actual pixel value may also be referred to as an original pixel value.

[0353] Optionally, in some possible implementations of the present invention, when the picture processing method is applied to a video decoding process, the method further includes: performing motion compensation decoding on the any pixel sample by using the motion vector that is of the any pixel sample in the current picture block and is obtained through computation, and obtaining a pixel reconstruction value of the any pixel sample.

[0354] Specifically, the process may be: determining the predicted pixel value of the pixel of the any pixel sample in the current picture block by using the motion vector that is of the any pixel sample in the current picture block and is obtained through computation; and performing motion compensation prediction on the any pixel sample by using the predicted pixel value of the pixel of the any pixel sample, so as to obtain a pixel reconstruction value of the any pixel sample.

[0355] Alternatively, the process may be: determining the predicted pixel value of the pixel of the any pixel sample in the current picture block by using the motion vector that is of the any pixel sample in the current picture block and is obtained through computation; and performing motion compensation prediction on the any pixel sample by using the predicted pixel value of the pixel of the any pixel sample, and decoding a bit stream to obtain a prediction residual of the any pixel sample, or decoding a bit stream to obtain a prediction residual of the current picture block, so as to obtain a prediction residual of the any pixel sample, and obtaining a pixel reconstruction value of the any pixel sample with reference to the pixel value that is of the pixel of the any pixel sample and is obtained by performing motion compensation prediction.

[0356] It may be understood that, picture processing may be performed on each picture block in a current video frame in a mode similar to a picture processing mode corresponding to the current picture block. Certainly, picture processing may be performed on some picture blocks in the current video frame in a mode different from the picture processing mode corresponding to the current picture block.

[0357] In the technical solution provided by this embodiment of the present invention, an affine motion model based on rotation and scaling motion is constructed by using only two parameters. This not only reduces computational complexity, but also improves motion vector estimation accuracy. After two displacement coefficients are introduced into the technical solution, motion vector estimation may be performed based on mixed motion of rotation, scaling, and translational motion according to the technical solution, so that motion vector estimation is more accurate.

[0358] For better understanding and implementing the foregoing solution of the embodiments of the present invention, the following uses specific application scenarios for further description.

picture block, in a video frame temporally adjacent to a video frame to which the current picture block belongs.

[0369] Optionally, in some possible implementations of the present invention, a candidate motion information unit set corresponding to the pixel sample in the lower region of the current picture block includes motion information units of $x7$ pixel samples. The $x7$ pixel samples include at least one pixel sample spatially adjacent to the pixel sample in the lower region of the current picture block and/or at least one pixel sample temporally adjacent to the pixel sample in the lower region of the current picture block, and $x7$ is a positive integer.

[0370] For example, $x7$ may be, for example, equal to 1, 2, 3, 4, 5, 6, or another value.

[0371] For example, the $x7$ pixel samples include at least one of a pixel sample that has a same location as the pixel sample in the lower region of the current picture block, a pixel sample spatially adjacent to a left edge of the current picture block, a pixel sample spatially adjacent to a lower left of the current picture block, or a pixel sample spatially adjacent to a lower edge of the current picture block, in a video frame temporally adjacent to the video frame to which the current picture block belongs.

[0372] Optionally, in some possible implementations of the present invention, a candidate motion information unit set corresponding to the pixel sample in the lower right region of the current picture block includes motion information units of $x8$ pixel samples. The $x8$ pixel samples include at least one pixel sample spatially adjacent to the pixel sample in the lower right region of the current picture block and/or at least one pixel sample temporally adjacent to the pixel sample in the lower right region of the current picture block, and $x8$ is a positive integer.

[0373] For example, $x8$ may be, for example, equal to 1, 2, 3, 4, 5, 6, or another value.

[0374] For example, the $x8$ pixel samples include at least one of a pixel sample that has a same location as the pixel sample in the lower right region of the current picture block, a pixel sample spatially adjacent to the right edge of the current picture block, a pixel sample spatially adjacent to a lower right of the current picture block, or a pixel sample spatially adjacent to the lower edge of the current picture block, in a video frame temporally adjacent to the video frame to which the current picture block belongs.

[0375] A candidate motion information unit set corresponding to a lower right pixel sample included in the pixel sample in the lower right region includes at least one pixel sample spatially adjacent to the lower right pixel sample of the current picture block and/or at least one pixel sample temporally adjacent to the lower right pixel sample of the current picture block, for example, may include at least one of a pixel sample that has a same location as the lower right pixel sample of the current picture block, a pixel sample spatially adjacent to the right edge of the current picture block,

a pixel sample spatially adjacent to the lower right of the current picture block, or a pixel sample spatially adjacent to the lower edge of the current picture block, in a video frame temporally adjacent to the video frame to which the current picture block belongs.

[0376] For candidate motion information unit sets corresponding to a lower left pixel sample, an upper right pixel sample, and a central pixel sample a_1 that are included in the pixel samples and methods for generating the corresponding candidate motion information unit sets, refer to the detailed description in the foregoing embodiment. Details are not described again herein.

[0377] Similarly, for methods for generating the candidate motion information unit sets corresponding to the pixel sample in the right region, the pixel sample in the lower region, the pixel sample in the lower right region, and the lower right pixel sample included in the pixel sample in the lower right region, refer to the method for generating the candidate motion information unit set corresponding to the lower left pixel sample, the upper right pixel sample, or the central pixel sample a_1 . Details are not described again herein.

[0378] S703. The video coding apparatus determines N candidate merged motion information unit sets based on the candidate motion information unit set corresponding to each of the two pixel samples.

[0379] For details of S703, refer to the detailed description in S203 in the foregoing embodiment. Details are not described again herein.

[0380] S704. The video coding apparatus determines, from the N candidate merged motion information unit sets, a merged motion information unit set i including two motion information units.

[0381] Optionally, in some possible implementations of the present invention, the video coding apparatus may further write an identifier of the merged motion information unit set i into a video bit stream. Correspondingly, a video decoding apparatus determines, from the N candidate merged motion information unit sets, based on the identifier that is of the merged motion information unit set i and is obtained from the video bit stream, the merged motion information unit set i including the two motion information units. The identifier of the merged motion information unit set i may be any information that can identify the merged motion information unit set i . For example, the identifier of the merged motion information unit set i may be an index of the merged motion information unit set i in a merged motion information unit set list.

[0382] In addition, for details of S704, refer to the detailed description in S204 in the foregoing embodiment. Details are not described again herein.

[0383] S705. The video coding apparatus obtains a motion vector 2-tuple by using the merged motion information unit set i .

[0384] Optionally, in some possible implementations of the present invention, the video coding apparatus may use two motion vectors in the merged information unit set i of the current picture block as motion vector predictors and as initial values to search for two motion vectors in the motion vector 2-tuple to perform a simplified affine motion search. The search process is described briefly as follows: using the motion vector predictors as initial values and performing an iterative update; and when a quantity of iterative updates reaches a specified quantity, or a matching error between a predicted value of the current picture block that is obtained according to two motion vectors obtained by performing an iterative update and an original value of the current picture block is less than a specified threshold, using the two motion vectors obtained by performing the iterative update as the motion vector 2-tuple.

[0385] Optionally, in some possible implementations of the present invention, the video coding apparatus may further obtain motion vector prediction residuals of the two pixel samples by using the two motion vectors in the merged information unit set i of the current picture block and the two motion vectors in the motion vector 2-tuple, namely, a difference between a motion vector corresponding to each motion vector in the merged information unit set i of the current picture block, in the motion vector 2-tuple and each motion vector in the merged information unit set i of the current picture block, and encode the motion vector prediction differences of the two pixel samples.

[0386] S706. The video coding apparatus obtains a motion vector of any pixel sample in the current picture block through computation by using an affine motion model and the motion vector 2-tuple.

[0387] The motion vector that is of any pixel sample in the current picture block and is obtained through computation may be any one of: the motion vector of each pixel in the current picture block, the motion vector of each pixel block in the current picture block, or the motion vector of any pixel sample in the current picture block in the foregoing embodiment. For details about the motion vector of each pixel in the current picture block, the motion vector of each pixel block in the current picture block, and the motion vector of any pixel sample in the current picture block in the foregoing embodiment, refer to the detailed description in the foregoing embodiment. Details are not described again herein.

[0388] The affine motion model may be in the following form:

$$\begin{cases} vx = ax + by \\ vy = -bx + ay \end{cases}, \text{ where}$$

(x, y) are coordinates of the any pixel sample, vx is a horizontal component of the motion vector of the any pixel sample, and vy is a vertical component of the motion vector of the

processing apparatus 1000 is applied to a video coding apparatus or the picture prediction apparatus is applied to a video decoding apparatus.

[0454] Optionally, in some possible implementations of the present invention, when the picture processing apparatus is applied to the video coding apparatus, the processor 1002 is further configured to: after obtaining the motion vector of the any pixel sample in the current picture block through computation by using the affine motion model and the motion vector 2-tuple, perform motion compensation predictive coding on the any pixel sample in the current picture block by using the motion vector that is of the any pixel sample in the current picture block and is obtained through computation.

[0455] Optionally, in some possible implementations of the present invention, when the picture processing apparatus is applied to the video decoding apparatus, the processor 1002 is further configured to: after determining a predicted pixel value of a pixel of the any pixel sample in the current picture block, perform motion compensation decoding on the any pixel sample by using the motion vector that is of the any pixel sample in the current picture block and is obtained through computation, and obtain a pixel reconstruction value of the any pixel sample.

[0456] It may be understood that, functions of functional units of the picture processing apparatus 1000 in this embodiment may be specifically implemented according to the method in the foregoing method embodiment. For a specific implementation process thereof, refer to the related description in the foregoing method embodiment. Details are not described again herein. The picture processing apparatus 1000 may be any apparatus that needs to output and play a video, for example, a device such as a notebook computer, a tablet computer, a personal computer, or a mobile phone.

[0457] In the technical solution provided by this embodiment of the present invention, a picture processing apparatus 1000 constructs an affine motion model based on rotation and scaling motion by using only two parameters. This not only reduces computational complexity, but also improves motion vector estimation accuracy. After two displacement coefficients are introduced into the picture processing apparatus 1000, the picture processing apparatus 1000 may perform motion vector estimation based on mixed motion of rotation, scaling, and translational motion, so that motion vector estimation is more accurate.

[0458] An embodiment of the present invention further provides a computer storage medium. The computer storage medium may store a program, and when the program is executed, at least some or all of the steps of any picture processing method in the foregoing method embodiments may be performed.

[0459] In the foregoing embodiments, the description of each embodiment has respective

focuses. For a part that is not described in detail in an embodiment, refer to related descriptions in other embodiments.

[0460] Referring to FIG. 11, FIG. 11 is a schematic flowchart of another picture processing method according to another embodiment of the present invention. As shown by way of example in FIG. 11, another picture processing method provided by another embodiment of the present invention may include the following steps.

[0461] S1101. Obtain coefficients of an affine motion model, and obtain a motion vector of any pixel sample in a current picture block through computation by using the coefficients of the affine motion model and the affine motion model.

[0462] Optionally, in some possible implementations of the present invention, the affine motion model may be in the following form:

$$\begin{cases} vx = ax + by \\ vy = -bx + ay \end{cases}, \text{ where}$$

(x, y) are coordinates of the any pixel sample, vx is a horizontal component of the motion vector of the any pixel sample, and vy is a vertical component of the motion vector of the any pixel sample; and

in the equation $vx = ax + by$, a is a horizontal coordinate coefficient of the horizontal component of the affine motion model, and b is a vertical coordinate coefficient of the horizontal

component of the affine motion model; and in the equation $vy = -bx + ay$, a is a vertical coordinate coefficient of the vertical component of the affine motion model, -b is a horizontal coordinate coefficient of the vertical component of the affine motion model, and the coefficients of the affine motion model may include a and b.

[0463] Optionally, in some possible implementations of the present invention, the coefficients of the affine motion model may further include a horizontal displacement coefficient c of the horizontal component of the affine motion model and a vertical displacement coefficient d of the vertical component of the affine motion model, and therefore, the affine motion model is in the following form:

$$\begin{cases} vx = ax + by + c \\ vy = -bx + ay + d \end{cases}$$

[0464] S1102. Determine a predicted pixel value of a pixel of the any pixel sample by using the motion vector that is of the any pixel sample and is obtained through computation.

horizontal component of the affine motion model and a vertical displacement coefficient d of the vertical component of the affine motion model, and therefore, the affine motion model is in the following form:

$$\begin{cases} vx = ax + by + c \\ vy = -bx + ay + d \end{cases}$$

5 **[0470]** For the detailed description in this embodiment, refer to the related description in the foregoing embodiment.

[0471] It may be understood that, functions of functional units of the picture processing apparatus 1200 in this embodiment may be specifically implemented according to the method in the foregoing method embodiment. For a specific implementation process thereof, refer to the related
10 description in the foregoing method embodiment. Details are not described again herein. The picture processing apparatus 1200 may be any apparatus that needs to output and play a video, for example, a device such as a notebook computer, a tablet computer, a personal computer, or a mobile phone.

[0472] In the technical solution provided by this embodiment of the present invention, the
15 picture processing apparatus 1200 constructs an affine motion model based on rotation and scaling motion by using only two parameters. This not only reduces computational complexity, but also improves motion vector estimation accuracy. After two displacement coefficients are introduced into the picture processing apparatus 1200, the picture processing apparatus 1200 may perform motion vector estimation based on mixed motion of rotation, scaling, and translational motion, so
20 that motion vector estimation is more accurate.

[0473] Referring to FIG. 13, FIG. 13 is a schematic diagram of a picture processing apparatus 1300 according to an embodiment of the present invention. The picture processing apparatus 1300 may include at least one bus 1301, at least one processor 1302 connected to the bus 1301, and at least one memory 1303 connected to the bus 1301.

25 **[0474]** The processor 1302 invokes, by using the bus 1301, code or an instruction stored in the memory 1303, so that the processor 1302 is configured to: obtain coefficients of an affine motion model, and obtain a motion vector of any pixel sample in the current picture block through computation by using the coefficients of the affine motion model and the affine motion model; and determine a predicted pixel value of a pixel of the any pixel sample by using the motion vector that
30 is of the any pixel sample and is obtained through computation.

[0475] Optionally, in some possible implementations of the present invention, the affine motion model may be in the following form:

$$\begin{cases} vx = ax + by \\ vy = -bx + ay \end{cases}, \text{ where}$$

(x, y) are coordinates of the any pixel sample, vx is a horizontal component of the motion vector of the any pixel sample, and vy is a vertical component of the motion vector of the any pixel sample; and

5 in the equation $vx = ax + by$, a is a horizontal coordinate coefficient of the horizontal component of the affine motion model, and b is a vertical coordinate coefficient of the horizontal

component of the affine motion model; and in the equation $vy = -bx + ay$, a is a vertical coordinate coefficient of the vertical component of the affine motion model, -b is a horizontal coordinate coefficient of the vertical component of the affine motion model, and the coefficients of
10 the affine motion model may include a and b.

[0476] Optionally, in some possible implementations of the present invention, the coefficients of the affine motion model may further include a horizontal displacement coefficient c of the horizontal component of the affine motion model and a vertical displacement coefficient d of the vertical component of the affine motion model, and therefore, the affine motion model is in the
15 following form:

$$\begin{cases} vx = ax + by + c \\ vy = -bx + ay + d \end{cases}$$

[0477] For the detailed description in this embodiment, refer to the related description in the foregoing embodiment.

20 [0478] It may be understood that, functions of functional units of the picture processing apparatus 1300 in this embodiment may be specifically implemented according to the method in the foregoing method embodiment. For a specific implementation process thereof, refer to the related description in the foregoing method embodiment. Details are not described again herein. The picture processing apparatus 1300 may be any apparatus that needs to output and play a video, for
25 example, a device such as a notebook computer, a tablet computer, a personal computer, or a mobile phone.

[0479] In the technical solution provided by this embodiment of the present invention, the picture processing apparatus 1300 constructs an affine motion model based on rotation and scaling motion by using only two parameters. This not only reduces computational complexity, but also
30 improves motion vector estimation accuracy. After two displacement coefficients are introduced

into the picture processing apparatus 1300, the picture processing apparatus 1300 may perform motion vector estimation based on mixed motion of rotation, scaling, and translational motion, so that motion vector estimation is more accurate.

[0480] An embodiment of the present invention further provides a computer storage medium.

5 The computer storage medium may store a program, and when the program is executed, at least some or all of the steps of any picture processing method in the foregoing method embodiments may be performed.

[0481] In the foregoing embodiments, the description of each embodiment has respective focuses. For a part that is not described in detail in an embodiment, refer to related descriptions in
10 other embodiments.

[0482] It should be noted that, to make the description brief, the foregoing method embodiments are expressed as a series of actions. However, persons skilled in the art should appreciate that the present invention is not limited to the described action sequence, because according to the present invention, some steps may be performed in other sequences or performed
15 simultaneously. In addition, persons skilled in the art should also appreciate that all the embodiments described in the specification are example embodiments, and the related actions and modules are not necessarily mandatory to the present invention.

[0483] In the several embodiments provided in the present application, it should be understood that the disclosed apparatus may be implemented in other manners. For example, the described
20 apparatus embodiment is merely an example. For example, the unit division is merely logical function division and may be other division in actual implementation. For example, a plurality of units or components may be combined or integrated into another system, or some features may be ignored or not performed. In addition, the displayed or discussed mutual couplings or direct couplings or communication connections may be implemented through some interfaces. The
25 indirect couplings or communication connections between the apparatuses or units may be implemented in electronic or other forms.

[0484] The units described as separate parts may or may not be physically separate, and parts displayed as units may or may not be physical units, may be located in one location, or may be distributed on a plurality of network units. Some or all of the units may be selected according to
30 actual requirements to achieve the objectives of the solutions of the embodiments.

[0485] In addition, functional units in the embodiments of the present invention may be integrated into one processing unit, or each of the units may exist alone physically, or two or more units are integrated into one unit. The integrated unit may be implemented in a form of hardware, or may be implemented in a form of a software functional unit.

[0486] When the foregoing integrated unit is implemented in the form of a software functional unit and sold or used as an independent product, the integrated unit may be stored in a computer-readable storage medium. Based on such an understanding, the technical solutions of the present invention essentially, or the part contributing to the prior art, or all or some of the technical solutions may be implemented in the form of a software product. The software product is stored in a storage medium and includes several instructions for instructing a computer device (which may be a personal computer, a server, or a network device, and may specifically be a processor in a computer device) to perform all or some of the steps of the foregoing methods described in the embodiments of the present invention. The foregoing storage medium includes: any medium that can store program code, such as a USB flash drive, a removable hard disk, a magnetic disk, an optical disc, a read-only memory (ROM, Read-Only Memory), or a random access memory (RAM, Random Access Memory).

[0487] The foregoing embodiments are merely intended for describing the technical solutions of the present invention, but not for limiting the present invention. Although the present invention is described in detail with reference to the foregoing embodiments, persons of ordinary skill in the art should understand that they may still make modifications to the technical solutions described in the foregoing embodiments or make equivalent replacements to some technical features thereof, without departing from the scope of the technical solutions of the embodiments of the present invention.

CLAIMS

1. A picture processing method, comprising:

obtaining a motion vector 2-tuple of a current picture block, wherein the motion vector 2-tuple comprises motion vectors of two pixel samples in a video frame to which the current picture block belongs; and

obtaining a motion vector of any pixel sample in the current picture block through computation by using an affine motion model and the motion vector 2-tuple; wherein

the affine motion model is in the following form:

$$\begin{cases} vx = ax + by \\ vy = -bx + ay \end{cases}, \text{ wherein}$$

(x, y) are coordinates of the any pixel sample, vx is a horizontal component of the motion vector of the any pixel sample, and vy is a vertical component of the motion vector of the any pixel sample; and

in the equation $vx = ax + by$, a is a horizontal coordinate coefficient of the horizontal component of the affine motion model, and b is a vertical coordinate coefficient of the horizontal

component of the affine motion model; and in the equation $vy = -bx + ay$, a is a vertical coordinate coefficient of the vertical component of the affine motion model, and -b is a horizontal coordinate coefficient of the vertical component of the affine motion model.

2. The method according to claim 1, wherein the affine motion model further comprises a horizontal displacement coefficient c of the horizontal component of the affine motion model and a vertical displacement coefficient d of the vertical component of the affine motion model, and therefore, the affine motion model is in the following form:

$$\begin{cases} vx = ax + by + c \\ vy = -bx + ay + d \end{cases}.$$

3. The method according to claim 1 or 2, wherein the obtaining a motion vector of any pixel sample in the current picture block through computation by using an affine motion model and the motion vector 2-tuple comprises:

obtaining values of the coefficients of the affine motion model by using the motion vectors of the two pixel samples and locations of the two pixel samples; and

obtaining the motion vector of the any pixel sample in the current picture block by using the affine motion model and the values of the coefficients of the affine motion model.

4. The method according to any one of claims 1 to 3, wherein the obtaining a motion vector of any pixel sample in the current picture block through computation by using an affine motion model and the motion vector 2-tuple comprises:

obtaining the values of the coefficients of the affine motion model by using a ratio of a difference between horizontal components of the motion vectors of the two pixel samples to a distance between the two pixel samples and a ratio of a difference between vertical components of the motion vectors of the two pixel samples to the distance between the two pixel samples; and

obtaining the motion vector of the any pixel sample in the current picture block by using the affine motion model and the values of the coefficients of the affine motion model.

5. The method according to any one of claims 1 to 3, wherein the obtaining a motion vector of any pixel sample in the current picture block through computation by using an affine motion model and the motion vector 2-tuple comprises:

obtaining the values of the coefficients of the affine motion model by using a ratio of a weighted sum of components of the motion vectors of the two pixel samples to a distance between the two pixel samples or to a square of a distance between the two pixel samples; and

obtaining the motion vector of the any pixel sample in the current picture block by using the affine motion model and the values of the coefficients of the affine motion model.

6. The method according to any one of claims 1 to 4, wherein when the two pixel samples comprise an upper left pixel sample of the current picture block and a pixel sample in a right region on a right side of the upper left pixel sample, the affine motion model is specifically:

$$\begin{cases} vx = \frac{vx_1 - vx_0}{w} x - \frac{vy_1 - vy_0}{w} y + vx_0 \\ vy = \frac{vy_1 - vy_0}{w} x + \frac{vx_1 - vx_0}{w} y + vy_0 \end{cases}, \text{ wherein}$$

(vx_0, vy_0) are a motion vector of the upper left pixel sample, (vx_1, vy_1) are a motion vector of the pixel sample in the right region, and w is a distance between the two pixel samples.

7. The method according to any one of claims 1 to 4, wherein when the two pixel samples comprise an upper left pixel sample of the current picture block and a pixel sample in a lower region below the upper left pixel sample, the affine motion model is specifically:

$$\begin{cases} vx = \frac{vy_2 - vy_0}{h} x + \frac{vx_2 - vx_0}{h} y + vx_0 \\ vy = -\frac{vx_2 - vx_0}{h} x + \frac{vy_2 - vy_0}{h} y + vy_0 \end{cases}, \text{ wherein}$$

(vx_0, vy_0) are a motion vector of the upper left pixel sample, (vx_2, vy_2) are a motion vector of the pixel sample in the lower region, and h is a distance between the two pixel samples.

8. The method according to any one of claim 1, 2, 3, or 5, wherein when the two pixel samples comprise an upper left pixel sample of the current picture block and a pixel sample in a lower right region on a lower right side of the upper left pixel sample, the affine motion model is specifically:

$$\begin{cases} vx = \frac{(vx_2w_1 + vy_2h_1) - (vx_0w_1 + vy_0h_1)}{w_1^2 + h_1^2}x + \frac{(vx_2h_1 - vy_2w_1) - (vx_0h_1 - vy_0w_1)}{w_1^2 + h_1^2}y + vx_0 \\ vy = -\frac{(vx_2h_1 - vy_2w_1) - (vx_0h_1 - vy_0w_1)}{w_1^2 + h_1^2}x + \frac{(vx_2w_1 + vy_2h_1) - (vx_0w_1 + vy_0h_1)}{w_1^2 + h_1^2}y + vy_0 \end{cases}, \text{ wherein}$$

(vx_0, vy_0) are a motion vector of the upper left pixel sample, (vx_2, vy_2) are a motion vector of the pixel sample in the lower right region, h_1 is a distance in a vertical direction between the two pixel samples, w_1 is a distance in a horizontal direction between the two pixel samples, and $w_1^2 + h_1^2$ is a square of the distance between the two pixel samples.

9. The method according to any one of claims 1 to 8, wherein after the obtaining a motion vector of any pixel sample in the current picture block through computation by using an affine motion model and the motion vector 2-tuple, the method further comprises:

performing motion compensation predictive coding on the any pixel sample in the current picture block by using the motion vector that is of the any pixel sample in the current picture block and is obtained through computation.

10. The method according to any one of claims 1 to 8, wherein after the obtaining a motion vector of any pixel sample in the current picture block through computation by using an affine motion model and the motion vector 2-tuple, the method further comprises:

performing motion compensation decoding on the any pixel sample by using the motion vector that is of the any pixel sample in the current picture block to obtain a pixel reconstruction value of the any pixel sample.

11. A picture processing apparatus, wherein the apparatus comprises:

an obtaining unit, configured to obtain a motion vector 2-tuple of a current picture block, wherein the motion vector 2-tuple comprises motion vectors of two pixel samples in a video frame to which the current picture block belongs; and

a computation unit, configured to obtain a motion vector of any pixel sample in the current picture block through computation by using an affine motion model and the motion vector 2-tuple obtained by the obtaining unit; wherein

the affine motion model is in the following form:

$$\begin{cases} vx = ax + by \\ vy = -bx + ay \end{cases}, \text{ wherein}$$

(x, y) are coordinates of the any pixel sample, vx is a horizontal component of the motion vector of the any pixel sample, and vy is a vertical component of the motion vector of the any pixel sample; and

in the equation $vx = ax + by$, a is a horizontal coordinate coefficient of the horizontal component of the affine motion model, and b is a vertical coordinate coefficient of the horizontal

component of the affine motion model; and in the equation $vy = -bx + ay$, a is a vertical coordinate coefficient of the vertical component of the affine motion model, and -b is a horizontal coordinate coefficient of the vertical component of the affine motion model.

12. The apparatus according to claim 11, wherein the affine motion model further comprises a horizontal displacement coefficient c of the horizontal component of the affine motion model and a vertical displacement coefficient d of the vertical component of the affine motion model, and therefore, the affine motion model is in the following form:

$$\begin{cases} vx = ax + by + c \\ vy = -bx + ay + d \end{cases}.$$

13. The apparatus according to claim 11 or 12, wherein the computation unit is specifically configured to:

obtain values of the coefficients of the affine motion model by using the motion vectors of the two pixel samples and locations of the two pixel samples; and

obtain the motion vector of the any pixel sample in the current picture block by using the affine motion model and the values of the coefficients of the affine motion model.

14. The apparatus according to any one of claims 11 to 13, wherein the computation unit is specifically configured to:

obtain the values of the coefficients of the affine motion model by using a ratio of a difference between horizontal components of the motion vectors of the two pixel samples to a distance between the two pixel samples and a ratio of a difference between vertical components of the motion vectors of the two pixel samples to the distance between the two pixel samples; and

obtain the motion vector of the any pixel sample in the current picture block by using the affine motion model and the values of the coefficients of the affine motion model.

15. The apparatus according to any one of claims 11 to 13, wherein the computation unit is specifically configured to:

obtain the values of the coefficients of the affine motion model by using a ratio of a weighted sum of components of the motion vectors of the two pixel samples to a distance between the two pixel samples or to a square of a distance between the two pixel samples; and

obtain the motion vector of the any pixel sample in the current picture block by using the affine motion model and the values of the coefficients of the affine motion model.

16. The apparatus according to any one of claims 11 to 14, wherein when the two pixel samples comprise an upper left pixel sample of the current picture block and a pixel sample in a right region on a right side of the upper left pixel sample, the affine motion model is specifically:

$$\begin{cases} vx = \frac{vx_1 - vx_0}{w} x - \frac{vy_1 - vy_0}{w} y + vx_0 \\ vy = \frac{vy_1 - vy_0}{w} x + \frac{vx_1 - vx_0}{w} y + vy_0 \end{cases}, \text{ wherein}$$

(vx_0, vy_0) are a motion vector of the upper left pixel sample, (vx_1, vy_1) are a motion vector of the pixel sample in the right region, and w is a distance between the two pixel samples.

17. The apparatus according to any one of claims 11 to 14, wherein when the two pixel samples comprise an upper left pixel sample of the current picture block and a pixel sample in a lower region below the upper left pixel sample, the affine motion model is specifically:

$$\begin{cases} vx = \frac{vy_2 - vy_0}{h} x + \frac{vx_2 - vx_0}{h} y + vx_0 \\ vy = -\frac{vx_2 - vx_0}{h} x + \frac{vy_2 - vy_0}{h} y + vy_0 \end{cases}, \text{ wherein}$$

(vx_0, vy_0) are a motion vector of the upper left pixel sample, (vx_2, vy_2) are a motion vector of the pixel sample in the lower region, and h is a distance between the two pixel samples.

18. The apparatus according to any one of claim 11, 12, 13, or 15, wherein when the two pixel samples comprise an upper left pixel sample of the current picture block and a pixel sample in a lower right region on a lower right side of the upper left pixel sample, the affine motion model is specifically:

$$\begin{cases} vx = \frac{(vx_3w_1 + vy_3h_1) - (vx_0w_1 + vy_0h_1)}{w_1^2 + h_1^2} x + \frac{(vx_3h_1 - vy_3w_1) - (vx_0h_1 - vy_0w_1)}{w_1^2 + h_1^2} y + vx_0 \\ vy = -\frac{(vx_3h_1 - vy_3w_1) - (vx_0h_1 - vy_0w_1)}{w_1^2 + h_1^2} x + \frac{(vx_3w_1 + vy_3h_1) - (vx_0w_1 + vy_0h_1)}{w_1^2 + h_1^2} y + vy_0 \end{cases}, \text{ wherein}$$

(vx_0, vy_0) are a motion vector of the upper left pixel sample, (vx_3, vy_3) are a motion vector of

the pixel sample in the lower right region, h_1 is a distance in a vertical direction between the two

pixel samples, w_1 is a distance in a horizontal direction between the two pixel samples, and $w_1^2 + h_1^2$ is a square of the distance between the two pixel samples.

19. The apparatus according to any one of claims 11 to 18, wherein when the picture processing apparatus is applied to a video coding apparatus, the apparatus further comprises an encoding unit, configured to perform motion compensation predictive coding on the any pixel sample in the current picture block by using the motion vector that is of the any pixel sample in the current picture block and is obtained by the computation unit through computation.

20. The apparatus according to any one of claims 11 to 18, wherein when the picture processing apparatus is applied to a video decoding apparatus, the apparatus further comprises a decoding unit, configured to perform motion compensation decoding on the any pixel sample by using the motion vector that is of the any pixel sample in the current picture block and is obtained by the computation unit through computation to obtain a pixel reconstruction value of the any pixel sample.

21. A picture processing method, comprising:

obtaining coefficients of an affine motion model, and obtaining a motion vector of any pixel sample in a current picture block through computation by using the coefficients of the affine motion model and the affine motion model; and

determining a predicted pixel value of a pixel of the any pixel sample by using the motion vector that is of the any pixel sample and is obtained through computation; wherein

the affine motion model is in the following form:

$$\begin{cases} vx = ax + by \\ vy = -bx + ay \end{cases}, \text{ wherein}$$

(x, y) are coordinates of the any pixel sample, vx is a horizontal component of the motion vector of the any pixel sample, and vy is a vertical component of the motion vector of the any pixel sample;

in the equation $vx = ax + by$, a is a horizontal coordinate coefficient of the horizontal component of the affine motion model, and b is a vertical coordinate coefficient of the horizontal

component of the affine motion model; and in the equation $vy = -bx + ay$, a is a vertical coordinate coefficient of the vertical component of the affine motion model, -b is a horizontal coordinate coefficient of the vertical component of the affine motion model, and the coefficients of

the affine motion model comprise a and b; and

the coefficients of the affine motion model further comprise a horizontal displacement coefficient c of the horizontal component of the affine motion model and a vertical displacement coefficient d of the vertical component of the affine motion model, and therefore, the affine motion model is in the following form:

$$\begin{cases} vx = ax + by + c \\ vy = -bx + ay + d \end{cases}$$

22. A picture processing apparatus, comprising:

an obtaining unit, configured to obtain coefficients of an affine motion model;

a computation unit, configured to obtain a motion vector of any pixel sample in the current picture block through computation by using the coefficients of the affine motion model that are obtained by the obtaining unit and the affine motion model; and

a predicting unit, configured to determine a predicted pixel value of a pixel of the any pixel sample by using the motion vector that is of the any pixel sample and is obtained by the computation unit through computation; wherein

the affine motion model is in the following form:

$$\begin{cases} vx = ax + by \\ vy = -bx + ay \end{cases}, \text{ wherein}$$

(x, y) are coordinates of the any pixel sample, vx is a horizontal component of the motion vector of the any pixel sample, and vy is a vertical component of the motion vector of the any pixel sample;

in the equation $vx = ax + by$, a is a horizontal coordinate coefficient of the horizontal component of the affine motion model, and b is a vertical coordinate coefficient of the horizontal

component of the affine motion model; and in the equation $vy = -bx + ay$, a is a vertical coordinate coefficient of the vertical component of the affine motion model, -b is a horizontal coordinate coefficient of the vertical component of the affine motion model, and the coefficients of the affine motion model comprise a and b; and

the coefficients of the affine motion model further comprise a horizontal displacement coefficient c of the horizontal component of the affine motion model and a vertical displacement coefficient d of the vertical component of the affine motion model, and therefore, the affine motion model is in the following form:

$$\begin{cases} vx = ax + by + c \\ vy = -bx + ay + d \end{cases}.$$