

MOTION VECTOR RANGE DERIVATION FOR ENHANCED INTERPOLATION FILTER

5

TECHNICAL FIELD

Embodiments of the present application (disclosure) generally relate to the field of picture
10 processing and more particularly to inter prediction, such as motion vector range derivation
for Enhanced interpolation filter.

BACKGROUND

15 Video coding (video encoding and decoding) is used in a wide range of digital video
applications, for example broadcast digital TV, video transmission over internet and mobile
networks, real-time conversational applications such as video chat, video conferencing, DVD
and Blu-ray discs, video content acquisition and editing systems, and camcorders of security
applications.

20 The amount of video data needed to depict even a relatively short video can be substantial,
which may result in difficulties when the data is to be streamed or otherwise communicated
across a communications network with limited bandwidth capacity. Thus, video data is
generally compressed before being communicated across modern day telecommunications
networks. The size of a video could also be an issue when the video is stored on a storage
25 device because memory resources may be limited. Video compression devices often use
software and/or hardware at the source to code the video data prior to transmission or storage,
thereby decreasing the quantity of data needed to represent digital video images. The
compressed data is then received at the destination by a video decompression device that
decodes the video data. With limited network resources and ever increasing demands of
30 higher video quality, improved compression and decompression techniques that improve
compression ratio with little to no sacrifice in picture quality are desirable.

In Essential Video Coding (EVC), in case of using affine prediction mode, pixel-based affine
transform motion compensation can be applied. Specifically, when applying an Enhanced bi-
linear Interpolation Filter (EIF), the filtering procedure comprises deriving a pixel-based

motion vector field from control point motion vectors, obtaining interpolated samples based on the derived motion vectors and filtering the result with a high-pass filter. In order to fulfill memory limitation requirements, a motion vector clipping may have to be applied. However, according to the prior art, a motion vector range is significantly reduced (sometimes to one
5 point) if the central motion vector is close to or at the picture boundary. Further, boundary motion vectors may not be clipped to a range corresponding to the EVC motion vector storage format.

SUMMARY

- 10 Aspects of the present application provide apparatuses and methods for encoding and decoding according to the independent claims.
- The foregoing and other objects are achieved by the subject matter of the independent claims. Further implementation forms are apparent from the dependent claims, the description and the figures.
- 15 Particular embodiments are outlined in the attached independent claims, with other embodiments in the dependent claims.

According to a first aspect, the disclosure relates to a method for coding video data, comprising the steps of: obtaining a center motion vector, `mv_center`, of a coding block,
20 deriving a first motion vector range for the coding block based on the center motion vector and a motion vector spread, wherein the motion vector spread depends on a size of the coding block; if the first motion vector range is at least partially pointing outside a first area including a reference picture, updating the first motion vector range to be pointing within the first area, such that a minimum value and/or a maximum value of the updated first motion
25 vector range is pointing at the boundary of the first area, wherein the difference between the maximum value and the minimum value of the updated first motion vector range is equal to the minimum of a double value of the motion vector spread and the size of the first area; and performing pixel-based motion compensation based on the updated first motion vector range.

The motion vector is a two-dimensional vector used for inter prediction that provides an
30 offset from the coordinates in the decoded picture to the coordinates in a reference picture. A reference picture contains samples that can be used for inter prediction in the decoding process of subsequent pictures in decoding order.

Defining $\text{horMax} - \text{horMin} = \text{Min}(\text{horMaxPic}, \text{horMinPic} + 2 * \text{deviationMV}[\log_2 \text{CbWidth} - 3]) - \text{horMinPic} = \text{Min}(\text{horMaxPic} - \text{horMinPic}, \text{horMinPic} + 2 * \text{deviationMV}[\log_2 \text{CbWidth} - 3] - \text{horMinPic}) = \text{Min}(\text{horMaxPic} - \text{horMinPic}, 2 * \text{deviationMV}[\log_2 \text{CbWidth} - 3])$, then

- 5 the difference between the maximum value and the minimum of the first MV horizontal component range of the updated first motion vector range may refer to $\text{horMax} - \text{horMin}$; the minimum of a double value of the horizontal motion vector spread and the width of the first area may refer to $\text{Min}(\text{horMaxPic} - \text{horMinPic}, 2 * \text{deviationMV}[\log_2 \text{CbWidth} - 3]) = \text{Min}(\text{the width of the first area, doubled horizontal motion vector spread})$, the width of the first area may refer to $\text{horMaxPic} - \text{horMinPic}$.

Similarly, defining $\text{verMax} - \text{verMin} = \text{Min}(\text{verMaxPic}, \text{verMinPic} + 2 * \text{deviationMV}[\log_2 \text{CbHeight} - 3]) - \text{verMinPic} = \text{Min}(\text{verMaxPic} - \text{verMinPic}, \text{verMinPic} + 2 * \text{deviationMV}[\log_2 \text{CbHeight} - 3] - \text{verMinPic}) = \text{Min}(\text{verMaxPic} - \text{verMinPic}, 2 * \text{deviationMV}[\log_2 \text{CbHeight} - 3])$,

- 15 then the difference between the maximum value and the minimum of the first MV vertical component range of the updated first motion vector range may refer to $\text{verMax} - \text{verMin}$; the minimum of a double value of the vertical motion vector spread and the height of the first area may refer to $\text{Min}(\text{verMaxPic} - \text{verMinPic}, 2 * \text{deviationMV}[\log_2 \text{CbHeight} - 3]) = \text{Min}(\text{the height of the first area, doubled vertical motion vector spread})$, the height of the first area may refer to $\text{verMaxPic} - \text{verMinPic}$.

Examples of the condition that the motion vector range is at least partially pointing outside a boundary of an area including a reference picture are shown in Fig. 12 and 13, in which the motion vector range (spread to the left and right of the central motion vector) is located partially outside the left boundary of the picture. Similar examples apply to the right, lower and upper boundaries.

The method according to the first aspect avoids a significant reduction of the motion vector range near the picture boundaries.

- According to an implementation form, the motion vector spread is represented by a horizontal motion vector spread and/or a vertical motion vector spread, and the horizontal motion vector spread is derived based on the width of the coding block, and the vertical motion vector spread is derived based on the height of the coding block.

According to another implementation form, the horizontal motion vector spread is represented by $\text{deviationMV}[\log_2 \text{CbWidth} - 3]$, and the vertical motion vector spread is

represented by $\text{deviationMV}[\log_2 \text{CbHeight} - 3]$, wherein cbWidth and cbHeight represent the width and the height of the coding block.

According to another implementation form, the array deviationMV is set equal to $\{128, 256, 544, 1120, 2272\}$.

- 5 According to another implementation form, the first motion vector range is represented by a first MV horizontal component range and/or a first MV vertical component range, the first MV horizontal component range comprises a first minimum MV horizontal component value, hor_min and a first maximum MV horizontal component value, hor_max , and the first MV vertical component range comprises a first minimum MV vertical component value, ver_min
10 and a first maximum MV vertical component value, ver_max .

According to another implementation form, the method further comprises deriving a second motion vector range based on a size of a picture comprising the coding block, wherein the second motion vector range is represented by a second MV horizontal component range and/or a second MV vertical component range, the second MV horizontal component range
15 comprises a second minimum MV horizontal component value, hor_min_pic and a second maximum MV horizontal component value, hor_max_pic and the second MV vertical component range comprises a second minimum MV vertical component value, ver_min_pic and a second maximum MV vertical component value, ver_max_pic .

- According to another implementation form, the deriving a second motion vector range based
20 on the size of a picture comprises: deriving a second motion vector range based on the size of a picture, a size of the extended area, the coding block position inside the picture and the size of the coding block.

- According to another implementation form, if the first motion vector range is at least partially pointing outside a first area including a reference picture, updating the first motion vector
25 range to be pointing within the first area, comprises: if the first minimum MV horizontal component value, hor_min is less than the second minimum MV horizontal component value, hor_min_pic , setting a updated value of the first minimum MV horizontal component value, hor_min equal to the second minimum MV horizontal component value, hor_min_pic , and deriving a updated value of the first maximum MV horizontal component value, hor_max
30 based on a sum of the second minimum MV horizontal component value, hor_min_pic and the double value of the horizontal motion vector spread.

Accordingly, the second minimum MV horizontal component value, hor_min_pic corresponds to the left boundary of the first area.

According to another implementation form, if the first motion vector range is at least partially pointing outside a first area including a reference picture, updating the first motion vector range to be pointing within the first area, comprises: if the first maximum MV horizontal component value, hor_max is greater than the second maximum MV horizontal component value, hor_max_pic, setting a updated value of the first maximum MV horizontal component value, hor_max equal to the second maximum MV horizontal component value, hor_max_pic, and deriving a updated value of the first minimum MV horizontal component value, hor_min based on a subtraction value of the second maximum MV horizontal component value, hor_max_pic and the doubled value of the horizontal motion vector spread.

Accordingly, the second maximum MV horizontal component value, hor_max_pic corresponds to the right boundary of the first area.

According to another implementation form, if the first motion vector range is at least partially pointing outside a first area including a reference picture, updating the first motion vector range to be pointing within the first area, comprises: if the first minimum MV vertical component value, ver_min is less than the second minimum MV vertical component value, ver_min_pic, setting a updated value of the first minimum MV vertical component value, ver_min equal to the second minimum MV vertical component value, ver_min_pic, and deriving a updated value of the first maximum MV vertical component value, ver_max, based on a sum of the second minimum MV vertical component value, ver_min_pic and the double value of the vertical motion vector spread.

Accordingly, the second minimum MV horizontal component value, ver_min_pic corresponds to the upper boundary of the first area.

According to another implementation form, if the first motion vector range is at least partially pointing outside a first area including a reference picture, updating the first motion vector range to be pointing within the first area, comprises: if the first maximum MV vertical component value, ver_max is greater than the second maximum MV vertical component value, ver_max_pic, setting a updated value of the first maximum MV vertical component value, ver_max equal to the second maximum MV vertical component value, ver_max_pic, and deriving a updated value of the first minimum MV vertical component value, ver_min, based on a subtraction value of the second maximum MV vertical component value, ver_max_pic and the doubled value of the vertical motion vector spread.

Accordingly, the second maximum MV vertical component value, ver_max_pic corresponds to the lower boundary of the first area.

According to another implementation form, if the first motion vector range is at least partially pointing outside a left boundary of the first area, updating the variables hor_min and hor_max, representing a first MV horizontal component range of the first motion vector range, as:

- 5 hor_min = hor_min_pic,
 hor_max = min (hor_max_pic, hor_min_pic + 2* horizontal motion vector spread);
 wherein hor_min indicates the updated first minimum MV horizontal component, and
 hor_max indicates the updated first maximum MV horizontal component; wherein
 hor_min_pic and hor_max_pic, representing a second MV horizontal component range of a
 10 second motion vector range; hor_min_pic indicates a second minimum MV horizontal
 component of the second MV horizontal component range, and hor_max_pic indicates a
 second maximum MV horizontal component of the second MV horizontal component range,
 and wherein the second motion vector range depends on a size of a picture comprising the
 coding block.
- 15 Here, if the first minimum MV horizontal component value, hor_min is smaller than the
 second minimum MV horizontal component value, and hor_min_pic corresponds to the left
 boundary of the first area.

According to another implementation form, if the first motion vector range is at least partially pointing outside a right boundary of the first area, updating the variables hor_min and

- 20 hor_max, representing a first MV horizontal component range of the first motion vector
 range, as:
 hor_min = max (hor_min_pic, hor_max_pic – 2* horizontal motion vector spread),
 hor_max = hor_max_pic;
 wherein hor_min indicates the updated first minimum MV horizontal component, and
 25 hor_max indicates the updated first maximum MV horizontal component; wherein
 hor_min_pic and hor_max_pic, representing a second MV horizontal component range of a
 second motion vector range; hor_min_pic indicates a second minimum MV horizontal
 component of the second MV horizontal component range, and hor_max_pic indicates a
 second maximum MV horizontal component of the second MV horizontal component range,
 30 and wherein the second motion vector range depends on a size of a picture comprising the
 coding block.
- Here, if the first maximum MV horizontal component value, hor_max is larger than the
 second maximum MV horizontal component value, and hor_max_pic corresponds to the right
 boundary of the first area.

According to another implementation form, if the first motion vector range is at least partially pointing outside an upper boundary of the first area, updating the variables `ver_min` and `ver_max`, representing a first MV vertical component range of the first motion vector range, as:

5 `ver_min = ver_min_pic,`
 `ver_max = min (ver_max_pic, ver_min_pic + 2* vertical motion vector spread);`
 wherein `ver_min` indicates the updated first minimum MV vertical component, and `ver_max` indicates the updated first maximum MV vertical component; wherein `ver_min_pic` and `ver_max_pic`, representing a second MV vertical component range of a second motion vector
 10 range; `ver_min_pic` indicates a second minimum MV vertical component of the second MV vertical component range, and `ver_max_pic` indicates a second maximum MV horizontal component of the second MV horizontal component range, and wherein the second motion vector range depends on a size of a picture comprising the coding block.

Here, if the first minimum MV vertical component value, `ver_min` is smaller than the second
 15 minimum MV horizontal component value, and `ver_min_pic` corresponds to the upper boundary of the first area.

According to another implementation form, if the first motion vector range is at least partially located outside a lower boundary of the first area, updating the variables `ver_min` and `ver_max`, representing a first MV vertical component range of the first motion vector range, as:

20 `ver_min = max (ver_min_pic, ver_max_pic - 2* vertical motion vector spread),`
 `ver_max = ver_max_pic;`
 wherein `ver_min` indicates the updated first minimum MV vertical component of the first MV vertical component range, and `ver_max` indicates the updated first maximum MV vertical
 25 component of the first MV vertical component range; wherein `ver_min_pic` and `ver_max_pic`, representing a second MV vertical component range of a second motion vector range;
 `ver_min_pic` indicates a second minimum MV vertical component of the second MV vertical component range, and `ver_max_pic` indicates a second maximum MV horizontal component
 30 of the second MV horizontal component range, and wherein the second motion vector range depends on a size of a picture comprising the coding block.

Here, if the first maximum MV vertical component value, `ver_max` is larger than the second maximum MV vertical component value, and `ver_max_pic` corresponds to the lower boundary of the first area. The size of the picture may be substantially the same as the size of the reference picture.

According to another implementation form, a minimum or maximum value of a horizontal component of the updated first motion vector range is pointing at a left or right boundary of the first area, respectively, and/or a minimum or maximum value of a vertical component of the updated first motion vector range is pointing at an upper or lower boundary of the first area, respectively.

According to another implementation form, the first area comprises the reference picture and an extended area surrounding the reference picture.

According to another implementation form, a size of the extended area depends on a coding tree unit, CTU, size. A margin may correspond to an area surrounding the reference picture, such that a margin size may correspond to the size of the extended area.

According to another implementation form, a size of the extended area is 128 pixels.

According to another implementation form, the method further comprises: performing a clipping operation on the updated first motion vector range to be within the range $[-2^{17}, 2^{17}-1]$; wherein the performing pixel-based motion compensation based on the updated first motion vector range, comprises: performing pixel-based motion compensation based on the updated and clipped first motion vector range.

According to another implementation form, the variables hor_min, ver_min, hor_max and ver_max represent correspondingly the first minimum MV horizontal component value, the first minimum MV vertical component value, the first maximum MV horizontal component value, and the first maximum MV vertical component value of the updated first motion vector range are clipped as follows:

hor_max = Clip3($-2^{17}, 2^{17}-1$, hor_max),
 ver_max = Clip3($-2^{17}, 2^{17}-1$, ver_max),
 hor_min = Clip3($-2^{17}, 2^{17}-1$, hor_min),
 ver_min = Clip3($-2^{17}, 2^{17}-1$, ver_min).

According to another implementation form, the performing pixel-based motion compensation based on the updated motion vector range, comprises: performing clipping operation on a motion vector of a pixel of the coding block to be within a range, to obtain a clipped motion vector, the range depends on the updated first motion vector range; and performing pixel-based motion compensation based on the clipped motion vector.

The motion vector of each pixel of the coding block is obtained based on the affine motion model, and the motion vector of a pixel of the extended area (margin area) surrounding the coding block is obtained based on the affine motion model.

According to another implementation form, obtaining a center motion vector of a coding block, is performed according to the following equations:

$$mv_center[0] = (mvBaseScaled[0] + dX[0] * (cbWidth >> 1) + dY[0] * (cbHeight >> 1)),$$

$$5 \quad mv_center[1] = (mvBaseScaled[1] + dX[1] * (cbWidth >> 1) + dY[1] * (cbHeight >> 1));$$

dX is a horizontal change of the motion vector according to the affine motion model,

dY is a vertical change of the motion vector, according to the affine motion model,

mvBaseScaled is a motion vector corresponding to the top left corner of the coding block,

10 according to the affine motion model for the coding block,

cbWidth and cbHeight are two variables specifying the width and the height of the coding block.

According to another implementation form, deriving a first motion vector range of the coding block based on the center motion vector and a motion vector spread is performed according

15 to the following equations:

$$hor_min = mv_center[0] - deviationMV[\log_2 CbWidth - 3],$$

$$ver_min = mv_center[1] - deviationMV[\log_2 CbHeight - 3],$$

$$hor_max = mv_center[0] + deviationMV[\log_2 CbWidth - 3],$$

$$ver_max = mv_center[1] + deviationMV[\log_2 CbHeight - 3];$$

20 wherein:

hor_min is the first minimum MV horizontal component value,

ver_min is the first minimum MV vertical component value,

hor_max is the first maximum MV horizontal component value,

ver_max is the first maximum MV vertical component value,

25 mv_center is the center motion vector,

wherein mv_center[0] corresponds to the horizontal and mv_center[1] corresponds to the vertical component of the center motion vector.

According to another implementation form, deriving a second motion vector range based on the size of the picture is performed according to the following equations:

$$30 \quad hor_max_pic = (pic_width + 128 - xCb - cbWidth - 1) << 5,$$

$$ver_max_pic = (pic_height + 128 - yCb - cbHeight - 1) << 5,$$

$$hor_min_pic = (-128 - xCb) << 5,$$

$$ver_min_pic = (-128 - yCb) << 5,$$

wherein:

- hor_min_pic is the second minimum MV horizontal component value,
ver_min_pic is the second minimum MV vertical component value,
hor_max_pic is the second maximum MV horizontal component value,
ver_max_pic is the second maximum MV vertical component value,
- 5 (xCb, yCb) is a location of the coding block in full-sample units,
cbWidth and cbHeight are two variables specifying the width and the height of the coding
block,
pic_width is a width of the picture in samples and
pic_height is a height of the picture in samples.
- 10 According to another implementation form, the pixel-based motion compensation is performed
using Enhanced Interpolation Filter (EIF). Accordingly, EIF may be used for an affine motion
model.
According to a second aspect, the disclosure relates to a coder comprising processing circuitry
for carrying out the method according to the first aspect or any of the implementation forms
15 thereof.
According to an implementation form, the coder comprises an encoder or a decoder.
According to a third aspect, the disclosure relates to a computer program product comprising
instructions which, when the program is executed by a computer, cause the computer to carry
out the method according to the first aspect or any of the implementation forms thereof.
- 20 According to a fourth aspect, the disclosure relates to a decoder, comprising: one or more
processors; and a non-transitory computer-readable storage medium coupled to the one or more
processors and storing instructions for execution by the one or more processors, wherein the
programming, when executed by the one or more processors, configures the decoder to carry
out the method according to the first aspect or any of the implementation forms thereof.
- 25 According to a fifth aspect, the disclosure relates to an encoder, comprising: one or more
processors; and a non-transitory computer-readable storage medium coupled to the one or more
processors and storing instructions for execution by the one or more processors, wherein the
programming, when executed by the processors, configures the encoder to carry out the method
according to the first aspect or any of the implementation forms thereof.
- 30 The method according to the first aspect of the invention can be performed by the coder
according to the second aspect of the invention. Further features and implementation forms of
the method according to the first aspect of the invention correspond to the features and
implementation forms of the coder according to the second aspect of the invention.

The method according to the first aspect of the invention can be performed by the decoder according to the fourth aspect of the invention. Further features and implementation forms of the method according to the first aspect of the invention correspond to the features and implementation forms of the decoder according to the fourth aspect of the invention.

- 5 The method according to the first aspect of the invention can be performed by the encoder according to the fifth aspect of the invention. Further features and implementation forms of the method according to the first aspect of the invention correspond to the features and implementation forms of the encoder according to the fifth aspect of the invention.

10 The advantages of the coder, decoder and encoder according to the third, fourth, and fifth aspect are the same as those for the method according to the first aspect and the corresponding implementation forms thereof.

Details of one or more embodiments are set forth in the accompanying drawings and the description below. Other features, objects, and advantages will be apparent from the description, drawings, and claims.

15

BRIEF DESCRIPTION OF THE DRAWINGS

In the following embodiments of the invention are described in more detail with reference to the attached figures and drawings, in which:

- FIG. 1A is a block diagram showing an example of a video coding system configured to
20 implement embodiments of the invention;
- FIG. 1B is a block diagram showing another example of a video coding system configured to implement embodiments of the invention;
- FIG. 2 is a block diagram showing an example of a video encoder configured to implement embodiments of the invention;
- 25 FIG. 3 is a block diagram showing an example structure of a video decoder configured to implement embodiments of the invention;
- FIG. 4 is a block diagram illustrating an example of an encoding apparatus or a decoding apparatus;
- FIG. 5 is a block diagram illustrating another example of an encoding apparatus or a
30 decoding apparatus;
- FIG. 6 is an illustration example of control point based affine motion model: 4-parameters and 6-parameters
- FIG. 7 is an illustration example of affine subblock motion vector field.

- FIG. 8 is an illustration example of the coordinates of corners of affine block (subblock) and of intermediate EIF block (subblock).
- FIG. 9 is an illustration example of the location of transformed block (subblock) in reference picture and corresponding bounding box.
- 5 FIG. 10 is an illustration example of motion vector range updating for case where central motion vector points outside the reference picture
- FIG. 11 is another illustration example of motion vector range updating, for case where central motion vector points outside the reference picture, with the real (uncropped) reference area depiction.
- 10 FIG. 12 is an illustration example of reducing motion vector range.
- FIG. 13 is another illustration example of reducing motion vector range.
- FIG. 14 illustrates a comparison between the conventional design and the design according to the present disclosure for case where motion vector range shrinks to one point near the picture boundaries for the conventional design.
- 15 FIG. 15 is a block diagram showing an example structure of a content supply system 3100 which realizes a content delivery service.
- FIG. 16 is a block diagram showing a structure of an example of a terminal device.
- FIG. 17 is a block diagram illustrating the method according to the present disclosure.
- FIG. 18 is a block diagram illustrating a decoder according to the present disclosure.
- 20 FIG. 19 is a block diagram illustrating an encoder according to the present disclosure.

In the following identical reference signs refer to identical or at least functionally equivalent features if not explicitly specified otherwise.

25 DETAILED DESCRIPTION OF THE EMBODIMENTS

In the following description, reference is made to the accompanying figures, which form part of the disclosure, and which show, by way of illustration, specific aspects of embodiments of the invention or specific aspects in which embodiments of the present invention may be used. It is understood that embodiments of the invention may be used in other aspects and comprise
30 structural or logical changes not depicted in the figures. The following detailed description, therefore, is not to be taken in a limiting sense, and the scope of the present invention is defined by the appended claims.

For instance, it is understood that a disclosure in connection with a described method may also hold true for a corresponding device or system configured to perform the method and

vice versa. For example, if one or a plurality of specific method steps are described, a corresponding device may include one or a plurality of units, e.g. functional units, to perform the described one or plurality of method steps (e.g. one unit performing the one or plurality of steps, or a plurality of units each performing one or more of the plurality of steps), even if
5 such one or more units are not explicitly described or illustrated in the figures. On the other hand, for example, if a specific apparatus is described based on one or a plurality of units, e.g. functional units, a corresponding method may include one step to perform the functionality of the one or plurality of units (e.g. one step performing the functionality of the one or plurality of units, or a plurality of steps each performing the functionality of one or more of the
10 plurality of units), even if such one or plurality of steps are not explicitly described or illustrated in the figures. Further, it is understood that the features of the various exemplary embodiments and/or aspects described herein may be combined with each other, unless specifically noted otherwise.

Video coding typically refers to the processing of a sequence of pictures, which form the video or video sequence. Instead of the term “picture” the term “frame” or “image” may be
15 used as synonyms in the field of video coding. Video coding (or coding in general) comprises two parts video encoding and video decoding. Video encoding is performed at the source side, typically comprising processing (e.g. by compression) the original video pictures to reduce the amount of data required for representing the video pictures (for more efficient
20 storage and/or transmission). Video decoding is performed at the destination side and typically comprises the inverse processing compared to the encoder to reconstruct the video pictures. Embodiments referring to “coding” of video pictures (or pictures in general) shall be understood to relate to “encoding” or “decoding” of video pictures or respective video sequences. The combination of the encoding part and the decoding part is also referred to as
25 CODEC (Coding and Decoding).

In case of lossless video coding, the original video pictures can be reconstructed, i.e. the reconstructed video pictures have the same quality as the original video pictures (assuming no transmission loss or other data loss during storage or transmission). In case of lossy video coding, further compression, e.g. by quantization, is performed, to reduce the amount of data
30 representing the video pictures, which cannot be completely reconstructed at the decoder, i.e. the quality of the reconstructed video pictures is lower or worse compared to the quality of the original video pictures.

Several video coding standards belong to the group of “lossy hybrid video codecs” (i.e. combine spatial and temporal prediction in the sample domain and 2D transform coding for

applying quantization in the transform domain). Each picture of a video sequence is typically partitioned into a set of non-overlapping blocks and the coding is typically performed on a block level. In other words, at the encoder the video is typically processed, i.e. encoded, on a block (video block) level, e.g. by using spatial (intra picture) prediction and/or temporal (inter picture) prediction to generate a prediction block, subtracting the prediction block from the current block (block currently processed/to be processed) to obtain a residual block, transforming the residual block and quantizing the residual block in the transform domain to reduce the amount of data to be transmitted (compression), whereas at the decoder the inverse processing compared to the encoder is applied to the encoded or compressed block to reconstruct the current block for representation. Furthermore, the encoder duplicates the decoder processing loop such that both will generate identical predictions (e.g. intra- and inter predictions) and/or re-constructions for processing, i.e. coding, the subsequent blocks. In the following embodiments of a video coding system 10, a video encoder 20 and a video decoder 30 are described based on Figs. 1 to 3.

Fig. 1A is a schematic block diagram illustrating an example coding system 10, e.g. a video coding system 10 (or short coding system 10) that may utilize techniques of this present application. Video encoder 20 (or short encoder 20) and video decoder 30 (or short decoder 30) of video coding system 10 represent examples of devices that may be configured to perform techniques in accordance with various examples described in the present application.

As shown in FIG. 1A, the coding system 10 comprises a source device 12 configured to provide encoded picture data 21 e.g. to a destination device 14 for decoding the encoded picture data 13.

The source device 12 comprises an encoder 20, and may additionally, i.e. optionally, comprise a picture source 16, a pre-processor (or pre-processing unit) 18, e.g. a picture pre-processor 18, and a communication interface or communication unit 22.

The picture source 16 may comprise or be any kind of picture capturing device, for example a camera for capturing a real-world picture, and/or any kind of a picture generating device, for example a computer-graphics processor for generating a computer animated picture, or any kind of other device for obtaining and/or providing a real-world picture, a computer generated picture (e.g. a screen content, a virtual reality (VR) picture) and/or any combination thereof (e.g. an augmented reality (AR) picture). The picture source may be any kind of memory or storage storing any of the aforementioned pictures.

In distinction to the pre-processor 18 and the processing performed by the pre-processing unit 18, the picture or picture data 17 may also be referred to as raw picture or raw picture data 17.

5 Pre-processor 18 is configured to receive the (raw) picture data 17 and to perform pre-processing on the picture data 17 to obtain a pre-processed picture 19 or pre-processed picture data 19. Pre-processing performed by the pre-processor 18 may, e.g., comprise trimming, color format conversion (e.g. from RGB to YCbCr), color correction, or de-noising. It can be understood that the pre-processing unit 18 may be optional component.

10 The video encoder 20 is configured to receive the pre-processed picture data 19 and provide encoded picture data 21 (further details will be described below, e.g., based on Fig. 2).

Communication interface 22 of the source device 12 may be configured to receive the encoded picture data 21 and to transmit the encoded picture data 21 (or any further processed version thereof) over communication channel 13 to another device, e.g. the destination device 14 or any other device, for storage or direct reconstruction.

15 The destination device 14 comprises a decoder 30 (e.g. a video decoder 30), and may additionally, i.e. optionally, comprise a communication interface or communication unit 28, a post-processor 32 (or post-processing unit 32) and a display device 34.

20 The communication interface 28 of the destination device 14 is configured receive the encoded picture data 21 (or any further processed version thereof), e.g. directly from the source device 12 or from any other source, e.g. a storage device, e.g. an encoded picture data storage device, and provide the encoded picture data 21 to the decoder 30.

25 The communication interface 22 and the communication interface 28 may be configured to transmit or receive the encoded picture data 21 or encoded data 13 via a direct communication link between the source device 12 and the destination device 14, e.g. a direct wired or wireless connection, or via any kind of network, e.g. a wired or wireless network or any combination thereof, or any kind of private and public network, or any kind of combination thereof.

30 The communication interface 22 may be, e.g., configured to package the encoded picture data 21 into an appropriate format, e.g. packets, and/or process the encoded picture data using any kind of transmission encoding or processing for transmission over a communication link or communication network.

The communication interface 28, forming the counterpart of the communication interface 22, may be, e.g., configured to receive the transmitted data and process the transmission data

using any kind of corresponding transmission decoding or processing and/or de-packaging to obtain the encoded picture data 21.

Both, communication interface 22 and communication interface 28 may be configured as unidirectional communication interfaces as indicated by the arrow for the communication channel 13 in Fig. 1A pointing from the source device 12 to the destination device 14, or bi-directional communication interfaces, and may be configured, e.g. to send and receive messages, e.g. to set up a connection, to acknowledge and exchange any other information related to the communication link and/or data transmission, e.g. encoded picture data transmission.

10 The decoder 30 is configured to receive the encoded picture data 21 and provide decoded picture data 31 or a decoded picture 31 (further details will be described below, e.g., based on Fig. 3 or Fig. 5).

The post-processor 32 of destination device 14 is configured to post-process the decoded picture data 31 (also called reconstructed picture data), e.g. the decoded picture 31, to obtain post-processed picture data 33, e.g. a post-processed picture 33. The post-processing performed by the post-processing unit 32 may comprise, e.g. color format conversion (e.g. from YCbCr to RGB), color correction, trimming, or re-sampling, or any other processing, e.g. for preparing the decoded picture data 31 for display, e.g. by display device 34.

The display device 34 of the destination device 14 is configured to receive the post-processed picture data 33 for displaying the picture, e.g. to a user or viewer. The display device 34 may be or comprise any kind of display for representing the reconstructed picture, e.g. an integrated or external display or monitor. The displays may, e.g. comprise liquid crystal displays (LCD), organic light emitting diodes (OLED) displays, plasma displays, projectors, micro LED displays, liquid crystal on silicon (LCoS), digital light processor (DLP) or any kind of other display.

Although Fig. 1A depicts the source device 12 and the destination device 14 as separate devices, embodiments of devices may also comprise both or both functionalities, the source device 12 or corresponding functionality and the destination device 14 or corresponding functionality. In such embodiments the source device 12 or corresponding functionality and the destination device 14 or corresponding functionality may be implemented using the same hardware and/or software or by separate hardware and/or software or any combination thereof.

As will be apparent for the skilled person based on the description, the existence and (exact) split of functionalities of the different units or functionalities within the source device 12

and/or destination device 14 as shown in Fig. 1A may vary depending on the actual device and application.

The encoder **20** (e.g. a video encoder 20) or the decoder **30** (e.g. a video decoder 30) or both encoder 20 and decoder 30 may be implemented via processing circuitry as shown in Fig. 1B,

5 such as one or more microprocessors, digital signal processors (DSPs), application-specific integrated circuits (ASICs), field-programmable gate arrays (FPGAs), discrete logic, hardware, video coding dedicated or any combinations thereof. The encoder 20 may be implemented via processing circuitry 46 to embody the various modules as discussed with respect to encoder 20 of FIG. 2 and/or any other encoder system or subsystem described
10 herein. The decoder 30 may be implemented via processing circuitry 46 to embody the various modules as discussed with respect to decoder 30 of FIG. 3 and/or any other decoder system or subsystem described herein. The processing circuitry may be configured to perform the various operations as discussed later. As shown in fig. 5, if the techniques are implemented partially in software, a device may store instructions for the software in a
15 suitable, non-transitory computer-readable storage medium and may execute the instructions in hardware using one or more processors to perform the techniques of this disclosure. Either of video encoder **20** and video decoder **30** may be integrated as part of a combined encoder/decoder (CODEC) in a single device, for example, as shown in Fig. 1B.

Source device 12 and destination device 14 may comprise any of a wide range of devices,

20 including any kind of handheld or stationary devices, e.g. notebook or laptop computers, mobile phones, smart phones, tablets or tablet computers, cameras, desktop computers, set-top boxes, televisions, display devices, digital media players, video gaming consoles, video streaming devices (such as content services servers or content delivery servers), broadcast receiver device, broadcast transmitter device, or the like and may use no or any kind of
25 operating system. In some cases, the source device 12 and the destination device 14 may be equipped for wireless communication. Thus, the source device 12 and the destination device 14 may be wireless communication devices.

In some cases, video coding system **10** illustrated in Fig. 1A is merely an example and the techniques of the present application may apply to video coding settings (e.g., video encoding

30 or video decoding) that do not necessarily include any data communication between the encoding and decoding devices. In other examples, data is retrieved from a local memory, streamed over a network, or the like. A video encoding device may encode and store data to memory, and/or a video decoding device may retrieve and decode data from memory. In some examples, the encoding and decoding is performed by devices that do not communicate

with one another, but simply encode data to memory and/or retrieve and decode data from memory.

For convenience of description, embodiments of the invention are described herein, for example, by reference to High-Efficiency Video Coding (HEVC) or to the reference software of Versatile Video coding (VVC), the next generation video coding standard developed by the Joint Collaboration Team on Video Coding (JCT-VC) of ITU-T Video Coding Experts Group (VCEG) and ISO/IEC Motion Picture Experts Group (MPEG). One of ordinary skill in the art will understand that embodiments of the invention are not limited to HEVC or VVC.

Encoder and Encoding Method

Fig. 2 shows a schematic block diagram of an example video encoder 20 that is configured to implement the techniques of the present application. In the example of Fig. 2, the video encoder 20 comprises an input 201 (or input interface 201), a residual calculation unit 204, a transform processing unit 206, a quantization unit 208, an inverse quantization unit 210, and inverse transform processing unit 212, a reconstruction unit 214, a loop filter unit 220, a decoded picture buffer (DPB) 230, a mode selection unit 260, an entropy encoding unit 270 and an output 272 (or output interface 272). The mode selection unit 260 may include an inter prediction unit 244, an intra prediction unit 254 and a partitioning unit 262. Inter prediction unit 244 may include a motion estimation unit and a motion compensation unit (not shown). A video encoder 20 as shown in Fig. 2 may also be referred to as hybrid video encoder or a video encoder according to a hybrid video codec.

The residual calculation unit 204, the transform processing unit 206, the quantization unit 208, the mode selection unit 260 may be referred to as forming a forward signal path of the encoder 20, whereas the inverse quantization unit 210, the inverse transform processing unit 212, the reconstruction unit 214, the buffer 216, the loop filter 220, the decoded picture buffer (DPB) 230, the inter prediction unit 244 and the intra-prediction unit 254 may be referred to as forming a backward signal path of the video encoder 20, wherein the backward signal path of the video encoder 20 corresponds to the signal path of the decoder (see video decoder 30 in Fig. 3). The inverse quantization unit 210, the inverse transform processing unit 212, the reconstruction unit 214, the loop filter 220, the decoded picture buffer (DPB) 230, the inter prediction unit 244 and the intra-prediction unit 254 are also referred to forming the “built-in decoder” of video encoder 20.

Pictures & Picture Partitioning (Pictures & Blocks)

The encoder 20 may be configured to receive, e.g. via input 201, a picture 17 (or picture data 17), e.g. picture of a sequence of pictures forming a video or video sequence. The received

picture or picture data may also be a pre-processed picture 19 (or pre-processed picture data 19). For sake of simplicity the following description refers to the picture 17. The picture 17 may also be referred to as current picture or picture to be coded (in particular in video coding to distinguish the current picture from other pictures, e.g. previously encoded and/or decoded pictures of the same video sequence, i.e. the video sequence which also comprises the current picture).

A (digital) picture is or can be regarded as a two-dimensional array or matrix of samples with intensity values. A sample in the array may also be referred to as pixel (short form of picture element) or a pel. The number of samples in horizontal and vertical direction (or axis) of the array or picture define the size and/or resolution of the picture. For representation of color, typically three color components are employed, i.e. the picture may be represented or include three sample arrays. In RGB format or color space a picture comprises a corresponding red, green and blue sample array. However, in video coding each pixel is typically represented in a luminance and chrominance format or color space, e.g. YCbCr, which comprises a luminance component indicated by Y (sometimes also L is used instead) and two chrominance components indicated by Cb and Cr. The luminance (or short luma) component Y represents the brightness or grey level intensity (e.g. like in a grey-scale picture), while the two chrominance (or short chroma) components Cb and Cr represent the chromaticity or color information components. Accordingly, a picture in YCbCr format comprises a luminance sample array of luminance sample values (Y), and two chrominance sample arrays of chrominance values (Cb and Cr). Pictures in RGB format may be converted or transformed into YCbCr format and vice versa, the process is also known as color transformation or conversion. If a picture is monochrome, the picture may comprise only a luminance sample array. Accordingly, a picture may be, for example, an array of luma samples in monochrome format or an array of luma samples and two corresponding arrays of chroma samples in 4:2:0, 4:2:2, and 4:4:4 colour format.

Embodiments of the video encoder 20 may comprise a picture partitioning unit (not depicted in Fig. 2) configured to partition the picture 17 into a plurality of (typically non-overlapping) picture blocks 203. These blocks may also be referred to as root blocks, macro blocks (H.264/AVC) or coding tree blocks (CTB) or coding tree units (CTU) (H.265/HEVC and VVC). The picture partitioning unit may be configured to use the same block size for all pictures of a video sequence and the corresponding grid defining the block size, or to change the block size between pictures or subsets or groups of pictures, and partition each picture into the corresponding blocks.

In further embodiments, the video encoder may be configured to receive directly a block 203 of the picture 17, e.g. one, several or all blocks forming the picture 17. The picture block 203 may also be referred to as current picture block or picture block to be coded.

Like the picture 17, the picture block 203 again is or can be regarded as a two-dimensional
5 array or matrix of samples with intensity values (sample values), although of smaller dimension than the picture 17. In other words, the block 203 may comprise, e.g., one sample array (e.g. a luma array in case of a monochrome picture 17, or a luma or chroma array in case of a color picture) or three sample arrays (e.g. a luma and two chroma arrays in case of a color picture 17) or any other number and/or kind of arrays depending on the color format
10 applied. The number of samples in horizontal and vertical direction (or axis) of the block 203 define the size of block 203. Accordingly, a block may, for example, an $M \times N$ (M -column by N -row) array of samples, or an $M \times N$ array of transform coefficients.

Embodiments of the video encoder 20 as shown in Fig. 2 may be configured to encode the picture 17 block by block, e.g. the encoding and prediction is performed per block 203.

15 Embodiments of the video encoder 20 as shown in Fig. 2 may be further configured to partition and/or encode the picture by using slices (also referred to as video slices), wherein a picture may be partitioned into or encoded using one or more slices (typically non-overlapping), and each slice may comprise one or more blocks (e.g. CTUs) or one or more groups of blocks (e.g. tiles (H.265/HEVC and VVC) or bricks (VVC)).

20 Embodiments of the video encoder 20 as shown in Fig. 2 may be further configured to partition and/or encode the picture by using slices/tile groups (also referred to as video tile groups) and/or tiles (also referred to as video tiles), wherein a picture may be partitioned into or encoded using one or more slices/tile groups (typically non-overlapping), and each slice/tile group may comprise, e.g. one or more blocks (e.g. CTUs) or one or more tiles,
25 wherein each tile, e.g. may be of rectangular shape and may comprise one or more blocks (e.g. CTUs), e.g. complete or fractional blocks.

Residual Calculation

The residual calculation unit 204 may be configured to calculate a residual block 205 (also referred to as residual 205) based on the picture block 203 and a prediction block 265 (further
30 details about the prediction block 265 are provided later), e.g. by subtracting sample values of the prediction block 265 from sample values of the picture block 203, sample by sample (pixel by pixel) to obtain the residual block 205 in the sample domain.

Transform

The transform processing unit 206 may be configured to apply a transform, e.g. a discrete cosine transform (DCT) or discrete sine transform (DST), on the sample values of the residual block 205 to obtain transform coefficients 207 in a transform domain. The transform coefficients 207 may also be referred to as transform residual coefficients and represent the residual block 205 in the transform domain.

The transform processing unit 206 may be configured to apply integer approximations of DCT/DST, such as the transforms specified for H.265/HEVC. Compared to an orthogonal DCT transform, such integer approximations are typically scaled by a certain factor. In order to preserve the norm of the residual block which is processed by forward and inverse transforms, additional scaling factors are applied as part of the transform process. The scaling factors are typically chosen based on certain constraints like scaling factors being a power of two for shift operations, bit depth of the transform coefficients, tradeoff between accuracy and implementation costs, etc. Specific scaling factors are, for example, specified for the inverse transform, e.g. by inverse transform processing unit 212 (and the corresponding inverse transform, e.g. by inverse transform processing unit 312 at video decoder 30) and corresponding scaling factors for the forward transform, e.g. by transform processing unit 206, at an encoder 20 may be specified accordingly.

Embodiments of the video encoder 20 (respectively transform processing unit 206) may be configured to output transform parameters, e.g. a type of transform or transforms, e.g. directly or encoded or compressed via the entropy encoding unit 270, so that, e.g., the video decoder 30 may receive and use the transform parameters for decoding.

Quantization

The quantization unit 208 may be configured to quantize the transform coefficients 207 to obtain quantized coefficients 209, e.g. by applying scalar quantization or vector quantization. The quantized coefficients 209 may also be referred to as quantized transform coefficients 209 or quantized residual coefficients 209.

The quantization process may reduce the bit depth associated with some or all of the transform coefficients 207. For example, an n-bit transform coefficient may be rounded down to an m-bit Transform coefficient during quantization, where n is greater than m. The degree of quantization may be modified by adjusting a quantization parameter (QP). For example for scalar quantization, different scaling may be applied to achieve finer or coarser quantization. Smaller quantization step sizes correspond to finer quantization, whereas larger quantization step sizes correspond to coarser quantization. The applicable quantization step size may be indicated by a quantization parameter (QP). The quantization parameter may for example be

an index to a predefined set of applicable quantization step sizes. For example, small quantization parameters may correspond to fine quantization (small quantization step sizes) and large quantization parameters may correspond to coarse quantization (large quantization step sizes) or vice versa. The quantization may include division by a quantization step size and a corresponding and/or the inverse dequantization, e.g. by inverse quantization unit 210, may include multiplication by the quantization step size. Embodiments according to some standards, e.g. HEVC, may be configured to use a quantization parameter to determine the quantization step size. Generally, the quantization step size may be calculated based on a quantization parameter using a fixed point approximation of an equation including division.

Additional scaling factors may be introduced for quantization and dequantization to restore the norm of the residual block, which might get modified because of the scaling used in the fixed point approximation of the equation for quantization step size and quantization parameter. In one example implementation, the scaling of the inverse transform and dequantization might be combined. Alternatively, customized quantization tables may be used and signaled from an encoder to a decoder, e.g. in a bitstream. The quantization is a lossy operation, wherein the loss increases with increasing quantization step sizes.

Embodiments of the video encoder 20 (respectively quantization unit 208) may be configured to output quantization parameters (QP), e.g. directly or encoded via the entropy encoding unit 270, so that, e.g., the video decoder 30 may receive and apply the quantization parameters for decoding.

Inverse Quantization

The inverse quantization unit 210 is configured to apply the inverse quantization of the quantization unit 208 on the quantized coefficients to obtain dequantized coefficients 211, e.g. by applying the inverse of the quantization scheme applied by the quantization unit 208 based on or using the same quantization step size as the quantization unit 208. The dequantized coefficients 211 may also be referred to as dequantized residual coefficients 211 and correspond - although typically not identical to the transform coefficients due to the loss by quantization - to the transform coefficients 207.

Inverse Transform

The inverse transform processing unit 212 is configured to apply the inverse transform of the transform applied by the transform processing unit 206, e.g. an inverse discrete cosine transform (DCT) or inverse discrete sine transform (DST) or other inverse transforms, to obtain a reconstructed residual block 213 (or corresponding dequantized coefficients 213)

in the sample domain. The reconstructed residual block 213 may also be referred to as transform block 213.

Reconstruction

The reconstruction unit 214 (e.g. adder or summer 214) is configured to add the transform block 213 (i.e. reconstructed residual block 213) to the prediction block 265 to obtain a reconstructed block 215 in the sample domain, e.g. by adding – sample by sample - the sample values of the reconstructed residual block 213 and the sample values of the prediction block 265.

Filtering

The loop filter unit 220 (or short “loop filter” 220), is configured to filter the reconstructed block 215 to obtain a filtered block 221, or in general, to filter reconstructed samples to obtain filtered sample values. The loop filter unit is, e.g., configured to smooth pixel transitions, or otherwise improve the video quality. The loop filter unit 220 may comprise one or more loop filters such as a de-blocking filter, a sample-adaptive offset (SAO) filter or one or more other filters, e.g. an adaptive loop filter (ALF), a noise suppression filter (NSF), or any combination thereof. In an example, the loop filter unit 220 may comprise a de-blocking filter, a SAO filter and an ALF filter. The order of the filtering process may be the deblocking filter, SAO and ALF. In another example, a process called the luma mapping with chroma scaling (LMCS) (namely, the adaptive in-loop reshaper) is added. This process is performed before deblocking. In another example, the deblocking filter process may be also applied to internal sub-block edges, e.g. affine sub-blocks edges, ATMVP sub-blocks edges, sub-block transform (SBT) edges and intra sub-partition (ISP) edges. Although the loop filter unit 220 is shown in FIG. 2 as being an in loop filter, in other configurations, the loop filter unit 220 may be implemented as a post loop filter. The filtered block 221 may also be referred to as filtered reconstructed block 221.

Embodiments of the video encoder 20 (respectively loop filter unit 220) may be configured to output loop filter parameters (such as SAO filter parameters or ALF filter parameters or LMCS parameters), e.g. directly or encoded via the entropy encoding unit 270, so that, e.g., a decoder 30 may receive and apply the same loop filter parameters or respective loop filters for decoding.

Decoded Picture Buffer

The decoded picture buffer (DPB) 230 may be a memory that stores reference pictures, or in general reference picture data, for encoding video data by video encoder 20. The DPB 230 may be formed by any of a variety of memory devices, such as dynamic random access

memory (DRAM), including synchronous DRAM (SDRAM), magnetoresistive RAM (MRAM), resistive RAM (RRAM), or other types of memory devices. The decoded picture buffer (DPB) 230 may be configured to store one or more filtered blocks 221. The decoded picture buffer 230 may be further configured to store other previously filtered blocks, e.g. previously reconstructed and filtered blocks 221, of the same current picture or of different pictures, e.g. previously reconstructed pictures, and may provide complete previously reconstructed, i.e. decoded, pictures (and corresponding reference blocks and samples) and/or a partially reconstructed current picture (and corresponding reference blocks and samples), for example for inter prediction. The decoded picture buffer (DPB) 230 may be also configured to store one or more unfiltered reconstructed blocks 215, or in general unfiltered reconstructed samples, e.g. if the reconstructed block 215 is not filtered by loop filter unit 220, or any other further processed version of the reconstructed blocks or samples.

Mode Selection (Partitioning & Prediction)

The mode selection unit 260 comprises partitioning unit 262, inter-prediction unit 244 and intra-prediction unit 254, and is configured to receive or obtain original picture data, e.g. an original block 203 (current block 203 of the current picture 17), and reconstructed picture data, e.g. filtered and/or unfiltered reconstructed samples or blocks of the same (current) picture and/or from one or a plurality of previously decoded pictures, e.g. from decoded picture buffer 230 or other buffers (e.g. line buffer, not shown).. The reconstructed picture data is used as reference picture data for prediction, e.g. inter-prediction or intra-prediction, to obtain a prediction block 265 or predictor 265.

Mode selection unit 260 may be configured to determine or select a partitioning for a current block prediction mode (including no partitioning) and a prediction mode (e.g. an intra or inter prediction mode) and generate a corresponding prediction block 265, which is used for the calculation of the residual block 205 and for the reconstruction of the reconstructed block 215.

Embodiments of the mode selection unit 260 may be configured to select the partitioning and the prediction mode (e.g. from those supported by or available for mode selection unit 260), which provide the best match or in other words the minimum residual (minimum residual means better compression for transmission or storage), or a minimum signaling overhead (minimum signaling overhead means better compression for transmission or storage), or which considers or balances both. The mode selection unit 260 may be configured to determine the partitioning and prediction mode based on rate distortion optimization (RDO), i.e. select the prediction mode which provides a minimum rate distortion. Terms like “best”,

“minimum”, “optimum” etc. in this context do not necessarily refer to an overall “best”, “minimum”, “optimum”, etc. but may also refer to the fulfillment of a termination or selection criterion like a value exceeding or falling below a threshold or other constraints leading potentially to a “sub-optimum selection” but reducing complexity and processing
5 time.

In other words, the partitioning unit 262 may be configured to partition a picture from a video sequence into a sequence of coding tree units (CTUs), and the CTU 203 may be further partitioned into smaller block partitions or sub-blocks (which form again blocks), e.g. iteratively using quad-tree-partitioning (QT), binary partitioning (BT) or triple-tree-
10 partitioning (TT) or any combination thereof, and to perform, e.g., the prediction for each of the block partitions or sub-blocks, wherein the mode selection comprises the selection of the tree-structure of the partitioned block 203 and the prediction modes are applied to each of the block partitions or sub-blocks.

In the following the partitioning (e.g. by partitioning unit 260) and prediction processing (by
15 inter-prediction unit 244 and intra-prediction unit 254) performed by an example video encoder 20 will be explained in more detail.

Partitioning

The partitioning unit 262 may be configured to partition a picture from a video sequence into a sequence of coding tree units (CTUs), and the partitioning unit 262 may partition (or split) a
20 coding tree unit (CTU) 203 into smaller partitions, e.g. smaller blocks of square or rectangular size. For a picture that has three sample arrays, a CTU consists of an $N \times N$ block of luma samples together with two corresponding blocks of chroma samples. The maximum allowed size of the luma block in a CTU is specified to be 128×128 in the developing versatile video coding (VVC), but it can be specified to be value rather than 128×128 in the
25 future, for example, 256×256 . The CTUs of a picture may be clustered/grouped as slices/tile groups, tiles or bricks. A tile covers a rectangular region of a picture, and a tile can be divided into one or more bricks. A brick consists of a number of CTU rows within a tile. A tile that is not partitioned into multiple bricks can be referred to as a brick. However, a brick is a true subset of a tile and is not referred to as a tile. There are two modes of tile groups are
30 supported in VVC, namely the raster-scan slice/tile group mode and the rectangular slice mode. In the raster-scan tile group mode, a slice/tile group contains a sequence of tiles in tile raster scan of a picture. In the rectangular slice mode, a slice contains a number of bricks of a picture that collectively form a rectangular region of the picture. The bricks within a rectangular slice are in the order of brick raster scan of the slice. These smaller blocks (which

may also be referred to as sub-blocks) may be further partitioned into even smaller partitions. This is also referred to as tree-partitioning or hierarchical tree-partitioning, wherein a root block, e.g. at root tree-level 0 (hierarchy-level 0, depth 0), may be recursively partitioned, e.g. partitioned into two or more blocks of a next lower tree-level, e.g. nodes at tree-level 1 (hierarchy-level 1, depth 1), wherein these blocks may be again partitioned into two or more blocks of a next lower level, e.g. tree-level 2 (hierarchy-level 2, depth 2), etc. until the partitioning is terminated, e.g. because a termination criterion is fulfilled, e.g. a maximum tree depth or minimum block size is reached. Blocks which are not further partitioned are also referred to as leaf-blocks or leaf nodes of the tree. A tree using partitioning into two partitions is referred to as binary-tree (BT), a tree using partitioning into three partitions is referred to as ternary-tree (TT), and a tree using partitioning into four partitions is referred to as quad-tree (QT).

For example, a coding tree unit (CTU) may be or comprise a CTB of luma samples, two corresponding CTBs of chroma samples of a picture that has three sample arrays, or a CTB of samples of a monochrome picture or a picture that is coded using three separate colour planes and syntax structures used to code the samples. Correspondingly, a coding tree block (CTB) may be an $N \times N$ block of samples for some value of N such that the division of a component into CTBs is a partitioning. A coding unit (CU) may be or comprise a coding block of luma samples, two corresponding coding blocks of chroma samples of a picture that has three sample arrays, or a coding block of samples of a monochrome picture or a picture that is coded using three separate colour planes and syntax structures used to code the samples. Correspondingly a coding block (CB) may be an $M \times N$ block of samples for some values of M and N such that the division of a CTB into coding blocks is a partitioning.

In embodiments, e.g., according to HEVC, a coding tree unit (CTU) may be split into CUs by using a quad-tree structure denoted as coding tree. The decision whether to code a picture area using inter-picture (temporal) or intra-picture (spatial) prediction is made at the leaf CU level. Each leaf CU can be further split into one, two or four PUs according to the PU splitting type. Inside one PU, the same prediction process is applied and the relevant information is transmitted to the decoder on a PU basis. After obtaining the residual block by applying the prediction process based on the PU splitting type, a leaf CU can be partitioned into transform units (TUs) according to another quadtree structure similar to the coding tree for the CU.

In embodiments, e.g., according to the latest video coding standard currently in development, which is referred to as Versatile Video Coding (VVC), a combined Quad-tree nested multi-type tree using binary and ternary splits segmentation structure for example used to partition a coding tree unit. In the coding tree structure within a coding tree unit, a CU can have either

5 a square or rectangular shape. For example, the coding tree unit (CTU) is first partitioned by a quaternary tree. Then the quaternary tree leaf nodes can be further partitioned by a multi-type tree structure. There are four splitting types in multi-type tree structure, vertical binary splitting (SPLIT_BT_VER), horizontal binary splitting (SPLIT_BT_HOR), vertical ternary splitting (SPLIT_TT_VER), and horizontal ternary splitting (SPLIT_TT_HOR). The multi-

10 type tree leaf nodes are called coding units (CUs), and unless the CU is too large for the maximum transform length, this segmentation is used for prediction and transform processing without any further partitioning. This means that, in most cases, the CU, PU and TU have the same block size in the quadtree with nested multi-type tree coding block structure. The exception occurs when maximum supported transform length is smaller than the width or

15 height of the colour component of the CU. VVC develops a unique signaling mechanism of the partition splitting information in quadtree with nested multi-type tree coding tree structure. In the signalling mechanism, a coding tree unit (CTU) is treated as the root of a quaternary tree and is first partitioned by a quaternary tree structure. Each quaternary tree leaf node (when sufficiently large to allow it) is then further partitioned by a multi-type tree

20 structure. In the multi-type tree structure, a first flag (mtt_split_cu_flag) is signalled to indicate whether the node is further partitioned; when a node is further partitioned, a second flag (mtt_split_cu_vertical_flag) is signalled to indicate the splitting direction, and then a third flag (mtt_split_cu_binary_flag) is signalled to indicate whether the split is a binary split or a ternary split. Based on the values of mtt_split_cu_vertical_flag and

25 mtt_split_cu_binary_flag, the multi-type tree splitting mode (MttSplitMode) of a CU can be derived by a decoder based on a predefined rule or a table. It should be noted, for a certain design, for example, 64×64 Luma block and 32×32 Chroma pipelining design in VVC hardware decoders, TT split is forbidden when either width or height of a luma coding block is larger than 64, as shown in Figure 6. TT split is also forbidden when either width or height

30 of a chroma coding block is larger than 32. The pipelining design will divide a picture into Virtual pipeline data units (VPDUs) which are defined as non-overlapping units in a picture. In hardware decoders, successive VPDUs are processed by multiple pipeline stages simultaneously. The VPDU size is roughly proportional to the buffer size in most pipeline stages, so it is important to keep the VPDU size small. In most hardware decoders, the VPDU

size can be set to maximum transform block (TB) size. However, in VVC, ternary tree (TT) and binary tree (BT) partition may lead to the increasing of VPDUs size.s

In addition, it should be noted that, when a portion of a tree node block exceeds the bottom or right picture boundary, the tree node block is forced to be split until the all samples of every
5 coded CU are located inside the picture boundaries.

As an example, the Intra Sub-Partitions (ISP) tool may divide luma intra-predicted blocks vertically or horizontally into 2 or 4 sub-partitions depending on the block size.

In one example, the mode selection unit 260 of video encoder 20 may be configured to perform any combination of the partitioning techniques described herein.

10 As described above, the video encoder 20 is configured to determine or select the best or an optimum prediction mode from a set of (e.g. pre-determined) prediction modes. The set of prediction modes may comprise, e.g., intra-prediction modes and/or inter-prediction modes.

Intra-Prediction

The set of intra-prediction modes may comprise 35 different intra-prediction modes, e.g. non-
15 directional modes like DC (or mean) mode and planar mode, or directional modes, e.g. as defined in HEVC, or may comprise 67 different intra-prediction modes, e.g. non-directional modes like DC (or mean) mode and planar mode, or directional modes, e.g. as defined for VVC. As an example, several conventional angular intra prediction modes are adaptively replaced with wide-angle intra prediction modes for the non-square blocks, e.g. as defined in
20 VVC. As another example, to avoid division operations for DC prediction, only the longer side is used to compute the average for non-square blocks. And, the results of intra prediction of planar mode may be further modified by a position dependent intra prediction combination (PDPC) method.

The intra-prediction unit 254 is configured to use reconstructed samples of neighboring
25 blocks of the same current picture to generate an intra-prediction block 265 according to an intra-prediction mode of the set of intra-prediction modes.

The intra prediction unit 254 (or in general the mode selection unit 260) is further configured to output intra-prediction parameters (or in general information indicative of the selected intra prediction mode for the block) to the entropy encoding unit 270 in form of syntax

30 elements 266 for inclusion into the encoded picture data 21, so that, e.g., the video decoder 30 may receive and use the prediction parameters for decoding.

Inter-Prediction

The set of (or possible) inter-prediction modes depends on the available reference pictures (i.e. previous at least partially decoded pictures, e.g. stored in DBP 230) and other inter-

prediction parameters, e.g. whether the whole reference picture or only a part, e.g. a search window area around the area of the current block, of the reference picture is used for searching for a best matching reference block, and/or e.g. whether pixel interpolation is applied, e.g. half/semi-pel, quarter-pel and/or 1/16 pel interpolation, or not.

- 5 Additional to the above prediction modes, skip mode, direct mode and/or other inter prediction mode may be applied.

For example, Extended merge prediction, the merge candidate list of such mode is constructed by including the following five types of candidates in order: Spatial MVP from spatial neighbor CUs, Temporal MVP from collocated CUs, History-based MVP from an

- 10 FIFO table, Pairwise average MVP and Zero MVs. And a bilateral-matching based decoder side motion vector refinement (DMVR) may be applied to increase the accuracy of the MVs of the merge mode. Merge mode with MVD (MMVD), which comes from merge mode with motion vector differences. A MMVD flag is signaled right after sending a skip flag and merge flag to specify whether MMVD mode is used for a CU. And a CU-level adaptive
- 15 motion vector resolution (AMVR) scheme may be applied. AMVR allows MVD of the CU to be coded in different precision. Dependent on the prediction mode for the current CU, the MVDs of the current CU can be adaptively selected. When a CU is coded in merge mode, the combined inter/intra prediction (CIIP) mode may be applied to the current CU. Weighted averaging of the inter and intra prediction signals is performed to obtain the CIIP prediction.
- 20 Affine motion compensated prediction, the affine motion field of the block is described by motion information of two control point (4-parameter) or three control point motion vectors (6-parameter). Subblock-based temporal motion vector prediction (SbTMVP), which is similar to the temporal motion vector prediction (TMVP) in HEVC, but predicts the motion vectors of the sub-CUs within the current CU. Bi-directional optical flow (BDOF), previously
- 25 referred to as BIO, is a simpler version that requires much less computation, especially in terms of number of multiplications and the size of the multiplier. Triangle partition mode, in such a mode, a CU is split evenly into two triangle-shaped partitions, using either the diagonal split or the anti-diagonal split. Besides, the bi-prediction mode is extended beyond simple averaging to allow weighted averaging of the two prediction signals.
- 30 The inter prediction unit 244 may include a motion estimation (ME) unit and a motion compensation (MC) unit (both not shown in Fig.2). The motion estimation unit may be configured to receive or obtain the picture block 203 (current picture block 203 of the current picture 17) and a decoded picture 231, or at least one or a plurality of previously reconstructed blocks, e.g. reconstructed blocks of one or a plurality of other/different

previously decoded pictures 231, for motion estimation. E.g. a video sequence may comprise the current picture and the previously decoded pictures 231, or in other words, the current picture and the previously decoded pictures 231 may be part of or form a sequence of pictures forming a video sequence.

- 5 The encoder 20 may, e.g., be configured to select a reference block from a plurality of reference blocks of the same or different pictures of the plurality of other pictures and provide a reference picture (or reference picture index) and/or an offset (spatial offset) between the position (x, y coordinates) of the reference block and the position of the current block as inter prediction parameters to the motion estimation unit. This offset is also called
10 motion vector (MV).

- The motion compensation unit is configured to obtain, e.g. receive, an inter prediction parameter and to perform inter prediction based on or using the inter prediction parameter to obtain an inter prediction block 265. Motion compensation, performed by the motion compensation unit, may involve fetching or generating the prediction block based on the
15 motion/block vector determined by motion estimation, possibly performing interpolations to sub-pixel precision. Interpolation filtering may generate additional pixel samples from known pixel samples, thus potentially increasing the number of candidate prediction blocks that may be used to code a picture block. Upon receiving the motion vector for the PU of the current picture block, the motion compensation unit may locate the prediction block to which the
20 motion vector points in one of the reference picture lists.

The motion compensation unit may also generate syntax elements associated with the blocks and video slices for use by video decoder 30 in decoding the picture blocks of the video slice. In addition or as an alternative to slices and respective syntax elements, tile groups and/or tiles and respective syntax elements may be generated or used.

25 Entropy Coding

- The entropy encoding unit 270 is configured to apply, for example, an entropy encoding algorithm or scheme (e.g. a variable length coding (VLC) scheme, an context adaptive VLC scheme (CAVLC), an arithmetic coding scheme, a binarization, a context adaptive binary arithmetic coding (CABAC), syntax-based context-adaptive binary arithmetic coding
30 (SBAC), probability interval partitioning entropy (PIPE) coding or another entropy encoding methodology or technique) or bypass (no compression) on the quantized coefficients 209, inter prediction parameters, intra prediction parameters, loop filter parameters and/or other syntax elements to obtain encoded picture data 21 which can be output via the output 272, e.g. in the form of an encoded bitstream 21, so that, e.g., the video decoder 30 may receive

and use the parameters for decoding, . The encoded bitstream 21 may be transmitted to video decoder 30, or stored in a memory for later transmission or retrieval by video decoder 30.

Other structural variations of the video encoder 20 can be used to encode the video stream.

For example, a non-transform based encoder 20 can quantize the residual signal directly

- 5 without the transform processing unit 206 for certain blocks or frames. In another implementation, an encoder 20 can have the quantization unit 208 and the inverse quantization unit 210 combined into a single unit.

Decoder and Decoding Method

Fig. 3 shows an example of a video decoder 30 that is configured to implement the

- 10 techniques of this present application. The video decoder 30 is configured to receive encoded picture data 21 (e.g. encoded bitstream 21), e.g. encoded by encoder 20, to obtain a decoded picture 331. The encoded picture data or bitstream comprises information for decoding the encoded picture data, e.g. data that represents picture blocks of an encoded video slice (and/or tile groups or tiles) and associated syntax elements.

- 15 In the example of Fig. 3, the decoder 30 comprises an entropy decoding unit 304, an inverse quantization unit 310, an inverse transform processing unit 312, a reconstruction unit 314 (e.g. a summer 314), a loop filter 320, a decoded picture buffer (DPB) 330, a mode application unit 360, an inter prediction unit 344 and an intra prediction unit 354. Inter prediction unit 344 may be or include a motion compensation unit. Video decoder 30 may, in
20 some examples, perform a decoding pass generally reciprocal to the encoding pass described with respect to video encoder 100 from FIG. 2.

As explained with regard to the encoder 20, the inverse quantization unit 210, the inverse transform processing unit 212, the reconstruction unit 214, the loop filter 220, the decoded picture buffer (DPB) 230, the inter prediction unit 344 and the intra prediction unit 354 are

- 25 also referred to as forming the “built-in decoder” of video encoder 20. Accordingly, the inverse quantization unit 310 may be identical in function to the inverse quantization unit 110, the inverse transform processing unit 312 may be identical in function to the inverse transform processing unit 212, the reconstruction unit 314 may be identical in function to reconstruction unit 214, the loop filter 320 may be identical in function to the loop filter 220,
30 and the decoded picture buffer 330 may be identical in function to the decoded picture buffer 230. Therefore, the explanations provided for the respective units and functions of the video 20 encoder apply correspondingly to the respective units and functions of the video decoder 30.

Entropy Decoding

- The entropy decoding unit 304 is configured to parse the bitstream 21 (or in general encoded picture data 21) and perform, for example, entropy decoding to the encoded picture data 21 to obtain, e.g., quantized coefficients 309 and/or decoded coding parameters (not shown in Fig. 3), e.g. any or all of inter prediction parameters (e.g. reference picture index and motion vector), intra prediction parameter (e.g. intra prediction mode or index), transform parameters, quantization parameters, loop filter parameters, and/or other syntax elements. Entropy decoding unit 304 maybe configured to apply the decoding algorithms or schemes corresponding to the encoding schemes as described with regard to the entropy encoding unit 270 of the encoder 20. Entropy decoding unit 304 may be further configured to provide inter prediction parameters, intra prediction parameter and/or other syntax elements to the mode application unit 360 and other parameters to other units of the decoder 30. Video decoder 30 may receive the syntax elements at the video slice level and/or the video block level. In addition or as an alternative to slices and respective syntax elements, tile groups and/or tiles and respective syntax elements may be received and/or used.
- 15 Inverse Quantization
- The inverse quantization unit 310 may be configured to receive quantization parameters (QP) (or in general information related to the inverse quantization) and quantized coefficients from the encoded picture data 21 (e.g. by parsing and/or decoding, e.g. by entropy decoding unit 304) and to apply based on the quantization parameters an inverse quantization on the
- 20 decoded quantized coefficients 309 to obtain dequantized coefficients 311, which may also be referred to as transform coefficients 311. The inverse quantization process may include use of a quantization parameter determined by video encoder 20 for each video block in the video slice (or tile or tile group) to determine a degree of quantization and, likewise, a degree of inverse quantization that should be applied.
- 25 Inverse Transform
- Inverse transform processing unit 312 may be configured to receive dequantized coefficients 311, also referred to as transform coefficients 311, and to apply a transform to the dequantized coefficients 311 in order to obtain reconstructed residual blocks 213 in the sample domain. The reconstructed residual blocks 213 may also be referred to as transform
- 30 blocks 313. The transform may be an inverse transform, e.g., an inverse DCT, an inverse DST, an inverse integer transform, or a conceptually similar inverse transform process. The inverse transform processing unit 312 may be further configured to receive transform parameters or corresponding information from the encoded picture data 21 (e.g. by parsing

and/or decoding, e.g. by entropy decoding unit 304) to determine the transform to be applied to the dequantized coefficients 311.

Reconstruction

The reconstruction unit 314 (e.g. adder or summer 314) may be configured to add the reconstructed residual block 313, to the prediction block 365 to obtain a reconstructed block 315 in the sample domain, e.g. by adding the sample values of the reconstructed residual block 313 and the sample values of the prediction block 365.

Filtering

The loop filter unit 320 (either in the coding loop or after the coding loop) is configured to filter the reconstructed block 315 to obtain a filtered block 321, e.g. to smooth pixel transitions, or otherwise improve the video quality. The loop filter unit 320 may comprise one or more loop filters such as a de-blocking filter, a sample-adaptive offset (SAO) filter or one or more other filters, e.g. an adaptive loop filter (ALF), a noise suppression filter (NSF), or any combination thereof. In an example, the loop filter unit 220 may comprise a de-blocking filter, a SAO filter and an ALF filter. The order of the filtering process may be the deblocking filter, SAO and ALF. In another example, a process called the luma mapping with chroma scaling (LMCS) (namely, the adaptive in-loop reshapener) is added. This process is performed before deblocking. In another example, the deblocking filter process may be also applied to internal sub-block edges, e.g. affine sub-blocks edges, ATMVP sub-blocks edges, sub-block transform (SBT) edges and intra sub-partition (ISP) edges. Although the loop filter unit 320 is shown in FIG. 3 as being an in loop filter, in other configurations, the loop filter unit 320 may be implemented as a post loop filter.

Decoded Picture Buffer

The decoded video blocks 321 of a picture are then stored in decoded picture buffer 330, which stores the decoded pictures 331 as reference pictures for subsequent motion compensation for other pictures and/or for output respectively display.

The decoder 30 is configured to output the decoded picture 311, e.g. via output 312, for presentation or viewing to a user.

Prediction

The inter prediction unit 344 may be identical to the inter prediction unit 244 (in particular to the motion compensation unit) and the intra prediction unit 354 may be identical to the inter prediction unit 254 in function, and performs split or partitioning decisions and prediction based on the partitioning and/or prediction parameters or respective information received from the encoded picture data 21 (e.g. by parsing and/or decoding, e.g. by entropy decoding

unit 304). Mode application unit 360 may be configured to perform the prediction (intra or inter prediction) per block based on reconstructed pictures, blocks or respective samples (filtered or unfiltered) to obtain the prediction block 365.

When the video slice is coded as an intra coded (I) slice, intra prediction unit 354 of mode application unit 360 is configured to generate prediction block 365 for a picture block of the current video slice based on a signaled intra prediction mode and data from previously decoded blocks of the current picture. When the video picture is coded as an inter coded (i.e., B, or P) slice, inter prediction unit 344 (e.g. motion compensation unit) of mode application unit 360 is configured to produce prediction blocks 365 for a video block of the current video slice based on the motion vectors and other syntax elements received from entropy decoding unit 304. For inter prediction, the prediction blocks may be produced from one of the reference pictures within one of the reference picture lists. Video decoder 30 may construct the reference frame lists, List 0 and List 1, using default construction techniques based on reference pictures stored in DPB 330. The same or similar may be applied for or by embodiments using tile groups (e.g. video tile groups) and/or tiles (e.g. video tiles) in addition or alternatively to slices (e.g. video slices), e.g. a video may be coded using I, P or B tile groups and /or tiles.

Mode application unit 360 is configured to determine the prediction information for a video block of the current video slice by parsing the motion vectors or related information and other syntax elements, and uses the prediction information to produce the prediction blocks for the current video block being decoded. For example, the mode application unit 360 uses some of the received syntax elements to determine a prediction mode (e.g., intra or inter prediction) used to code the video blocks of the video slice, an inter prediction slice type (e.g., B slice, P slice, or GPB slice), construction information for one or more of the reference picture lists for the slice, motion vectors for each inter encoded video block of the slice, inter prediction status for each inter coded video block of the slice, and other information to decode the video blocks in the current video slice. The same or similar may be applied for or by embodiments using tile groups (e.g. video tile groups) and/or tiles (e.g. video tiles) in addition or alternatively to slices (e.g. video slices), e.g. a video may be coded using I, P or B tile groups and/or tiles.

Embodiments of the video decoder 30 as shown in Fig. 3 may be configured to partition and/or decode the picture by using slices (also referred to as video slices), wherein a picture may be partitioned into or decoded using one or more slices (typically non-overlapping), and

each slice may comprise one or more blocks (e.g. CTUs) or one or more groups of blocks (e.g. tiles (H.265/HEVC and VVC) or bricks (VVC)).

Embodiments of the video decoder 30 as shown in Fig. 3 may be configured to partition and/or decode the picture by using slices/tile groups (also referred to as video tile groups) and/or tiles (also referred to as video tiles), wherein a picture may be partitioned into or decoded using one or more slices/tile groups (typically non-overlapping), and each slice/tile group may comprise, e.g. one or more blocks (e.g. CTUs) or one or more tiles, wherein each tile, e.g. may be of rectangular shape and may comprise one or more blocks (e.g. CTUs), e.g. complete or fractional blocks.

Other variations of the video decoder 30 can be used to decode the encoded picture data 21. For example, the decoder 30 can produce the output video stream without the loop filtering unit 320. For example, a non-transform based decoder 30 can inverse-quantize the residual signal directly without the inverse-transform processing unit 312 for certain blocks or frames. In another implementation, the video decoder 30 can have the inverse-quantization unit 310 and the inverse-transform processing unit 312 combined into a single unit.

It should be understood that, in the encoder 20 and the decoder 30, a processing result of a current step may be further processed and then output to the next step. For example, after interpolation filtering, motion vector derivation or loop filtering, a further operation, such as Clip or shift, may be performed on the processing result of the interpolation filtering, motion vector derivation or loop filtering.

It should be noted that further operations may be applied to the derived motion vectors of current block (including but not limit to control point motion vectors of affine mode, sub-block motion vectors in affine, planar, ATMVP modes, temporal motion vectors, and so on). For example, the value of motion vector is constrained to a predefined range according to its representing bit. If the representing bit of motion vector is bitDepth, then the range is $-2^{(bitDepth-1)} \sim 2^{(bitDepth-1)}-1$, where “^” means exponentiation. For example, if bitDepth is set equal to 16, the range is $-32768 \sim 32767$; if bitDepth is set equal to 18, the range is $-131072 \sim 131071$. For example, the value of the derived motion vector (e.g. the MVs of four 4x4 sub-blocks within one 8x8 block) is constrained such that the max difference between integer parts of the four 4x4 sub-block MVs is no more than N pixels, such as no more than 1 pixel. Here provides two methods for constraining the motion vector according to the bitDepth.

FIG. 4 is a schematic diagram of a video coding device 400 according to an embodiment of the disclosure. The video coding device 400 is suitable for implementing the disclosed

embodiments as described herein. In an embodiment, the video coding device 400 may be a decoder such as video decoder 30 of FIG. 1A or an encoder such as video encoder 20 of FIG. 1A.

The video coding device 400 comprises ingress ports 410 (or input ports 410) and receiver units (Rx) 420 for receiving data; a processor, logic unit, or central processing unit (CPU) 430 to process the data; transmitter units (Tx) 440 and egress ports 450 (or output ports 450) for transmitting the data; and a memory 460 for storing the data. The video coding device 400 may also comprise optical-to-electrical (OE) components and electrical-to-optical (EO) components coupled to the ingress ports 410, the receiver units 420, the transmitter units 440, and the egress ports 450 for egress or ingress of optical or electrical signals.

The processor 430 is implemented by hardware and software. The processor 430 may be implemented as one or more CPU chips, cores (e.g., as a multi-core processor), FPGAs, ASICs, and DSPs. The processor 430 is in communication with the ingress ports 410, receiver units 420, transmitter units 440, egress ports 450, and memory 460. The processor 430 comprises a coding module 470. The coding module 470 implements the disclosed embodiments described above. For instance, the coding module 470 implements, processes, prepares, or provides the various coding operations. The inclusion of the coding module 470 therefore provides a substantial improvement to the functionality of the video coding device 400 and effects a transformation of the video coding device 400 to a different state. Alternatively, the coding module 470 is implemented as instructions stored in the memory 460 and executed by the processor 430.

The memory 460 may comprise one or more disks, tape drives, and solid-state drives and may be used as an over-flow data storage device, to store programs when such programs are selected for execution, and to store instructions and data that are read during program execution. The memory 460 may be, for example, volatile and/or non-volatile and may be a read-only memory (ROM), random access memory (RAM), ternary content-addressable memory (TCAM), and/or static random-access memory (SRAM).

Fig. 5 is a simplified block diagram of an apparatus 500 that may be used as either or both of the source device 12 and the destination device 14 from Fig. 1 according to an exemplary embodiment.

A processor 502 in the apparatus 500 can be a central processing unit. Alternatively, the processor 502 can be any other type of device, or multiple devices, capable of manipulating or processing information now-existing or hereafter developed. Although the disclosed

implementations can be practiced with a single processor as shown, e.g., the processor 502, advantages in speed and efficiency can be achieved using more than one processor.

A memory 504 in the apparatus 500 can be a read only memory (ROM) device or a random access memory (RAM) device in an implementation. Any other suitable type of storage

5 device can be used as the memory 504. The memory 504 can include code and data 506 that is accessed by the processor 502 using a bus 512. The memory 504 can further include an operating system 508 and application programs 510, the application programs 510 including at least one program that permits the processor 502 to perform the methods described here.

For example, the application programs 510 can include applications 1 through N, which
10 further include a video coding application that performs the methods described here.

The apparatus 500 can also include one or more output devices, such as a display 518. The display 518 may be, in one example, a touch sensitive display that combines a display with a touch sensitive element that is operable to sense touch inputs. The display 518 can be coupled to the processor 502 via the bus 512.

15 Although depicted here as a single bus, the bus 512 of the apparatus 500 can be composed of multiple buses. Further, the secondary storage 514 can be directly coupled to the other components of the apparatus 500 or can be accessed via a network and can comprise a single integrated unit such as a memory card or multiple units such as multiple memory cards. The apparatus 500 can thus be implemented in a wide variety of configurations.

20 **1 Background**

In the following, background information for the present disclosure is presented.

1.1 Affine motion compensated prediction

In ITU-T H.265, only translation motion model is applied for motion compensation prediction (MCP). While in the real world, there are many kinds of motion, e.g. zoom in/out, rotation,
25 perspective motions and the other irregular motions. In the EVC, a block-based affine transform motion compensation prediction is applied. As shown at FIG. 6, the affine motion field of the block is described by motion information of two control point (4-parameter) or three control point motion vectors (CPMV) (6-parameter).

1.1.1 General equations for motion vector derivation

30 The general equation for calculation motion vector at sample location (x, y) is:

$$\begin{cases} mv_x = dHorX * x + dVerX * y + mv_{0x} \\ mv_y = dHorY * x + dVerY * y + mv_{0y} \end{cases} \quad (1-1)$$

For 4-parameter affine motion model, motion vector at sample location (x, y) is derived as:

$$\begin{cases} mv_x = \frac{mv_{1x}-mv_{0x}}{W} x + \frac{mv_{0y}-mv_{1y}}{W} y + mv_{0x} \\ mv_y = \frac{mv_{1y}-mv_{0y}}{W} x + \frac{mv_{1x}-mv_{0x}}{W} y + mv_{0y} \end{cases} \quad (1-2)$$

For 6-parameter affine motion model, motion vector at sample location (x, y) is derived as:

$$\begin{cases} mv_x = \frac{mv_{1x}-mv_{0x}}{W} x + \frac{mv_{2x}-mv_{0x}}{H} y + mv_{0x} \\ mv_y = \frac{mv_{1y}-mv_{0y}}{W} x + \frac{mv_{2y}-mv_{0y}}{H} y + mv_{0y} \end{cases} \quad (1-3)$$

Where (mv_{0x}, mv_{0y}) is motion vector of the top-left corner control point, (mv_{1x}, mv_{1y}) is motion vector of the top-right corner control point, and (mv_{2x}, mv_{2y}) is motion vector of the bottom-left corner control point.

For case of using 6-parameter affine motion model

$$dHorX = \frac{mv_{1x}-mv_{0x}}{W} \quad (1-4)$$

$$dHorY = \frac{mv_{1y}-mv_{0y}}{W}, \quad (1-5)$$

$$dVerX = \frac{mv_{2x}-mv_{0x}}{H}, \quad (1-6)$$

$$dVerY = \frac{mv_{2y}-mv_{0y}}{H} \quad (1-7)$$

For case of using 4-parameter affine motion model,

$$dHorX = \frac{mv_{1x}-mv_{0x}}{W}, \quad (1-8)$$

$$dHorY = \frac{mv_{1y}-mv_{0y}}{W}, \quad (1-9)$$

$$dVerX = -dHorY, \quad (1-10)$$

$$dVerY = dHorX. \quad (1-11)$$

As done for translational motion inter prediction, there are also two affine motion inter prediction modes: affine merge mode and affine AMVP mode.

1.1.2 Memory bandwidth calculation

Memory bandwidth is calculated as a reference block to current block area ratio. For example, for 8x8 bi-predictive block in case of usage interpolation filter with T taps, the reference area

value is S_r is equal to $2(8+T-1)(8+T-1)$, block area S_b is equal to $8*8$. So memory bandwidth is $MB_{8x8} = \frac{2(8+T-1)(8+T-1)}{64}$. For 8-tap DCTIF that is used in ITU-T H.265, VVC and EVC, $MB_{8x8} = \frac{2*15*15}{64} = 7.03125$.

5 1.2 Block based affine transform prediction

In order to simplify the motion compensation prediction, block based affine transform prediction is applied. To derive motion vector of each 8×8 luma sub-block, the motion vector of the center sample of each sub-block, as shown in FIG. 7, is calculated according to above equations, and rounded to $1/16$ fraction accuracy. Then the motion compensation interpolation filters are applied, to generate the prediction of each sub-block with derived motion vector. The sub-block size of chroma-components is set to be 4×4 .

1.2.1 Block based affine transform prediction with 8×8 subblocks

Subblock affine motion compensation with minimal subblock size 8×8 is much more hardware friendly than affine motion compensation with minimal subblock size 4×4 . There are at least three reasons for that.

1. Memory bandwidth. Affine motion compensation with minimal subblock size 8×8 does not increase the memory bandwidth compared to ITU-T H.265, because 8×8 bi-predictive blocks are the worst case for ITU-T H.265 in terms of memory bandwidth calculation. In EVC 8×8 bi-predictive blocks also does not change the worst case in terms of memory bandwidth ($8 \times 4/4 \times 8$ bi-predictive blocks are the worst case in EVC 3.0 and $4 \times 16/16 \times 4$ blocks are going to be the worst case in EVC 4.0). Basically 8×8 bi-predictive blocks can occur in regular inter prediction both in EVC and ITU-T H.265, so affine subblock motion compensation with such minimal block size does not increase complexity of motion compensation.
2. Number of multiplications. Motion compensation for 8×8 subblock requires much less number of multiplications than motion compensation of four 4×4 subblocks.
3. Memory access. In some hardware implementations no less than 16 samples can be read. From this perspective 8×8 blocks that in case of usage 8-tap DCTIF takes $(8+8-1)*(8+8-1)$ reference samples utilize memory much more efficient than 4×4 blocks.

However subblock affine motion compensation with minimal subblock size 8x8 is much provides significant performance drop in comparison with subblock affine motion compensation with minimal subblock size 4x4. Especially for content with fast rotation. For such content EIF can be used.

5 1.3 Enhanced bilinear interpolation filter

Enhanced bi-linear Interpolation Filter (EIF) can be used for a prediction block and on subblock basis. The filtering procedure is the same for luma and for chroma signals, the filtering procedure comprises of following steps:

1. Deriving pixel-based motion vector field from the CPMVs, according to the equation (1-1);
 2. Obtaining interpolated samples based on the derived motion vectors, using bilinear interpolation for the fractional offsets;
 3. Performing horizontal and then vertical filtering using fixed 3 - tap high-pass filter $[-1, 10, -1]$ with normalization factor of 8.
- First two steps are performed for $(w + 2) \times (h + 2)$ area, where w and h are prediction block width and height correspondently; one pixel margin from each border is added for applying 3 - tap filter at the third step. Original affine block and corresponding $(w + 2) \times (h + 2)$ block that is used in intermediate step of EIF are depicted in FIG. 8.

The example of EIF implementation is described below.

20 Interpolation process for the enhanced interpolation filter

Inputs to this process are:

- a location (x_{Cb} , y_{Cb}) in full-sample units,
- two variables $cbWidth$ and $cbHeight$ specifying the width and the height of the current coding block,
- horizontal change of motion vector dX ,
- vertical change of motion vector dY ,
- motion vector $mvBaseScaled$,
- the selected reference picture sample arrays $refPicLX$,
- sample bit depth $bitDepth$
- width of the picture in samples pic_width ,
- height of the picture in samples pic_height ,
- $clipMV$ flag specifying MV clipping type,
- $isLuma$ flag specifying whether luma or chroma is processed..

Outputs of this process are:

- an $(cbWidth / SubWidthC) \times (cbHeight / SubHeightC)$ array `predSamplesLX` of prediction sample values.

Interpolation filter coefficients $T[p]$ for each fractional sample position p equal to $xFrac$ or $yFrac$ are specified in Table 8-16.

- 5 The variables `hor_max`, `ver_max`, `hor_min` and `ver_min` are derived by invoking the process specified in 8.5.4.5 with a location (xCb, yCb) in full-sample units, two variables `cbWidth` and `cbHeight` specifying the width and the height of the current coding block, horizontal change of motion vector `dX`, vertical change of motion vector `dY`, motion vector `mvBaseScaled`, width of the picture in samples `pic_width`, height of the picture in samples `pic_height` and `clipMV` flag as input, and `hor_max`, `ver_max`, `hor_min` and `ver_min` as output.
- 10 If `isLuma` is equal to `FALSE`, then variables `mvBaseScaled`, `hor_min`, `hor_max`, `ver_min`, `ver_max` are modified as follows:
 - $xCb = xCb / SubWidthC$ (8-748)
 - $yCb = yCb / SubHeightC$ (8-749)
 - $cbWidth = cbWidth / SubWidthC$ (8-750)
 - 15 – $cbHeight = cbHeight / SubHeightC$ (8-751)
 - $mvBaseScaled[0] = mvBaseScaled[0] / SubWidthC$ (8-752)
 - $mvBaseScaled[1] = mvBaseScaled[1] / SubHeightC$ (8-753)
 - $hor_min = hor_min / SubWidthC$ (8-754)
 - $hor_max = hor_max / SubWidthC$ (8-755)
 - 20 – $ver_min = ver_min / SubHeightC$ (8-756)
 - $ver_max = ver_max / SubHeightC$ (8-757)

The variables `shift0`, `shift1`, `offset0` and `offset1` are derived as follows:

`shift0` is set equal to `bitDepth – 8`, `offset0` is equal to 0,

`shift1` is set equal to `12 – shift0`, `offset1` is equal to $2^{shift1-1}$

- 25 For $x = -1..cbWidth$ and $y = -1..cbHeight$, the following applies:
 - The motion vector `mvX` is derived as follows :
 - $mvX[0] = (mvBaseScaled[0] + dX[0] * x + dY[0] * y)$ (8-758)
 - $mvX[1] = (mvBaseScaled[1] + dX[1] * x + dY[1] * y)$ (8-759)
 - 30 $mvX[0] = Clip3(hor_min, hor_max, mvX[0])$ //a horizontal component of a clipped motion vector (8-760)
 - $mvX[1] = Clip3(ver_min, ver_max, mvX[1])$ //a vertical component of a clipped motion vector (8-761)
 - The variables `xInt`, `yInt`, `xFrac` and `yFrac` are derived as follows:
 - $xInt = xCb + (mvX[0] >> 9) + x$ (8-762)
 - 35 $yInt = yCb + (mvX[1] >> 9) + y$ (8-763)
 - $xFrac = (mvX[0] >> 4) \& 31$ (8-764)
 - $yFrac = (mvX[1] >> 4) \& 31$ (8-765)
 - The variables `A` and `B` are derived as follows:
 - $A = (refPicLX[xInt][yInt] * T[xFrac][0] +$

$$+ \text{refPicLX}[xInt + 1][yInt] * T[xFrac][1] + \text{offset0}) \gg \text{shift0} \quad (8-766)$$

$$B = (\text{refPicLX}[xInt][yInt + 1] * T[xFrac][0] +$$

$$+ \text{refPicLX}[xInt + 1][yInt + 1] * T[xFrac][1] + \text{offset0}) \gg \text{shift0} \quad (8-767)$$

– The sample value $b_{x,y}$ corresponding to location (x, y) is derived as follows:

$$b_{x,y} = (A * T[yFrac][0] + B * T[yFrac][1] + \text{offset1}) \gg \text{shift1} \quad (8-768)$$

The enhancement interpolation filter coefficients $eF[]$ are specified as $\{-1, 10, -1\}$.

The variables shift2 , shift3 , offset2 and offset3 are derived as follows:

shift2 is set equal to $\max(\text{bit_depth} - 11, 0)$, offset2 is equal to $2^{\text{shift2}-1}$,

shift3 is set equal to $(6 - \max(\text{bit_depth} - 11, 0))$, offset3 is equal to $2^{\text{shift3}-1}$,

10 For $x = 0.. \text{cbWidth} - 1$ and $y = -1.. \text{cbHeight}$, the following applies:

$$h_{x,y} = (eF[0] * b_{x-1,y} + eF[1] * b_{x,y} + eF[2] * b_{x+1,y} + \text{offset2}) \gg \text{shift2} \quad (8-769)$$

For $x = 0.. \text{cbWidth} - 1$ and $y = 0.. \text{cbHeight} - 1$, the following applies:

$$\begin{aligned} \text{predSamplesLXL}[x][y] &= \text{Clip3}(0, (1 \ll \text{bitDepth}) - 1, \\ & (eF[0] * h_{x,y-1} + eF[1] * h_{x,y} + eF[2] * b_{x,y+1} + \text{offset3}) \gg \text{shift3}) \end{aligned} \quad (8-770)$$

15

Table 8-16 – Specification of the interpolation filter coefficients $T[p]$ for each fractional sample position p

Fractional sample position p	Interpolation filter coefficients	
	$T[p][0]$	$T[p][1]$
0	64	0
1	62	2
2	60	4
3	58	6
4	56	8
5	54	10
6	52	12

7	50	14
8	48	16
9	46	18
10	44	20
11	42	22
12	40	24
13	38	26
14	36	28
15	34	30
16	32	32
17	30	34
18	28	36
19	26	38
20	24	40
21	22	42
22	20	44
23	18	46
24	16	48
25	14	50
26	12	52

27	10	54
28	8	56
29	6	58
30	4	60
31	2	62

1.4 Adaptive usage of EIF and block-based affine motion compensation.

EIF have less multiplications in comparison with subblock affine motion compensation with minimal subblock sizes 8x8 and 4x4. However memory bandwidth for EIF without affine motion model restrictions can be huge. Also for effective hardware implementations of EIF some additional requirements can appears. For example, the following requirements can appears for EIF from the hardware perspective.

- A. The internal buffer is limited by N lines, where N can be for example 3, 4, 5 or more. This means that during processing of one line (one row) of the current block (subblock) no more than N lines from the reference picture can be used.
- B. Memory access should be sequential, which means that if for i_{th} line of current block j_{th} line of the reference picture is fetched, then for $(i+1)_{th}$ line of the current block only lines $j+1, j+2, \dots$ can be fetched.
- C. No more than one additional line can be fetched for all lines of the current block except for the first.

To combine benefits of 8x8 subblock based affine motion compensation and EIF, adaptive scheme of usage EIF and subblock affine motion compensation.

1.4.1 Adaptive usage of EIF and block-based affine motion compensation. Basic algorithm.

Algorithm 1. Adaptive usage of EIF and block-based affine motion compensation (basic algorithm)

The basic algorithm of adaptive usage of EIF and subblock affine motion compensation is as follows.

1. Calculate optimal subblock size $M \times N$ based on affine motion model parameters.
2. If both optimal subblock width M and optimal subblock height N are greater or equal than
5 8, then perform subblock motion compensation with $M \times N$ subblock.
3. Otherwise, check EIF applicability conditions
 - 3.1. EIF applicability condition 1
 - 3.2. EIF applicability condition 2
 - 3.3. ...
 - 10 3.4. EIF applicability condition P
 - 3.5. If all EIF applicability conditions are met, then perform EIF motion compensation.
EIF motion compensation process comprise:
 - 3.5.1. Check EIF applicability condition $P+1$
 - 3.5.2. Check EIF applicability condition $P+2$
 - 15 3.5.3. ...
 - 3.5.4. Check EIF applicability condition $P+K$
 - 3.5.5. If one of EIF applicability conditions $P+1 \dots P+K$ are not met, then
 - 3.5.5.1. calculate first motion vector range that provide fulfilment of conditions
that were not met.
 - 20 3.5.5.2. set second motion vector range equal to the first motion vector range
 - 3.5.6. Otherwise
 - 3.5.6.1. calculate third motion vector range
 - 3.5.6.2. set second motion vector range equal to the third motion vector range
 - 3.5.7. Clip motion vectors calculated according to affine model to guarantee, that
25 these vectors are in the second motion vector range.
 - 3.6. Otherwise set $M = \max(M, 8)$, $N = \max(N, 8)$ and perform subblock motion
compensation with $M \times N$ subblock.

In some examples P and K can be equal to zero that means that the algorithm above will not
have steps 3.1 – 3.4 if P is equal to zero and will not have steps 3.5.1 – 3.5.4 if K is equal to
30 zero.

The details of some steps of this basic algorithm is described below.

1.4.1.1 Step 1. Optimal subblock size calculation

One method for deriving affine sub-block size is based on the motion vector differences of the affine control points and the width and height of an affine block. The sub-block size $M \times N$ could be derived as by Equation (2-1), where $MvPre$ is the motion vector accuracy (e.g., 1/4 pel in HEVC standard or 1/16 pel in VVC and EVC standards), and affine motion model parameters $dHorX$, $dHorY$, $dVerX$, $dVerY$ are calculated according to equations (1-4) – (1-7) for 6 parameter model and according to equations (1-8) – (1-11) for 4 parameter model.

$$\begin{cases} M = clip3(4, w, \frac{w \times MvPre}{\max(abs(dHorX), abs(dHorY))}) \\ N = clip3(4, h, \frac{h \times MvPre}{\max(abs(dVerX), abs(dVerY))}) \end{cases} \quad (2-1)$$

M and N derived in Equation (2-1) will be adjusted downward if necessary to make sure that w and h is divisible by M and N , respectively.

Another method is to build a 3 dimensional looking up table, then get the sub-block size directly from the looking up table according to motion vector difference, block size, and motion vector precision. For example, set M to $Table_M[x][y][z]$, where x equal to $\max(abs(v_{1x} - v_{0x}), abs(v_{1y} - v_{0y}))$, y equal to affine block width, z equal to the motion vector precision; set N to $Table_N[x][y][z]$, where x equal to $\max(abs(v_{2x} - v_{0x}), abs(v_{2y} - v_{0y}))$, y equal to affine block height, z equal to the motion vector precision.

In some exemplary implementation steps 1 is implemented as follows:

1.4.1.1.1 Derivation process for affine subblock size

Inputs to this process are:

- two variables $cbWidth$ and $cbHeight$ specifying the width and the height of the luma coding block,
- the number of control point motion vectors $numCpMv$,
- the control point motion vectors $cpMvLX[cpIdx]$, with $cpIdx = 0..numCpMv - 1$ and X being 0 or 1,
- the prediction list utilization flags $predFlagLX$, with X being 0 or 1.

Outputs of this process are:

- the size of luma coding subblocks in horizontal direction $sizeSbX$ and in vertical direction $sizeSbY$,

sizeSbX is set equal to cbWidth, sizeSbY is set equal to cbHeight.

When predFlagLX is equal to 1, the following applies for X being 0 and 1:

- Horizontal change of motion vector dX, vertical change of motion vector dY and base motion vector mvBaseScaled are derived by invoking the process specified in clause 1.4.1.1.2 with the luma coding block width cbWidth, the luma coding block height cbHeight, number of control point motion vectors numCpMv and the control point motion vectors cpMvLX[cpIdx] with cpIdx = 0..numCpMv – 1 as inputs.
- The variable mvWx and mvWy are derived as follows:
 - $mvWx = \max(\text{abs}(dX[0]), \text{abs}(dX[1]))$
 - $mvWy = \max(\text{abs}(dY[0]), \text{abs}(dY[1]))$
- The variable sizeSbXTemp is specified in Table 8-5 according to the value of mvWx.
- The variable sizeSbYTemp is specified in Table 8-5 according to the value of mvWy.
- The variable sizeSbX is modified as follow:
 - $sizeSbX = \min(sizeSbX, sizeSbXTemp)$
- The variable sizeSbY is modified as follow:
 - $sizeSbY = \min(sizeSbY, sizeSbYTemp)$

Table 8-5 – Specification of sizeSbXTemp for various input values of mvWx

mvWx	0	1	2	3	4	> 4
sizeSbX	cbWidth	32	16	8	8	4

Table 8-5 – Specification of sizeSbYTemp for various input values of mvWy

mvWx	0	1	2	3	4	> 4
sizeSbY	cbHeight	32	16	8	8	4

- 20 Where clause 1.4.1.1.2 is described as follows:

1.4.1.1.2 Derivation process for affine motion model parameters from control point motion vectors

Inputs to this process are:

- two variables cbWidth and cbHeight specifying the width and the height of the luma coding block,
- 25 – the number of control point motion vectors numCpMv,

- the control point motion vectors $cpMvLX[cpIdx]$, with $cpIdx = 0..numCpMv - 1$ and X being 0 or 1.

Outputs of this process are:

- horizontal change of motion vector dX ,
 - vertical change of motion vector dY ,
- 5 – motion vector $mvBaseScaled$ corresponding to the top left corner of the luma coding block.

The variables $\log2CbW$ and $\log2CbH$ are derived as follows:

$$\log2CbW = \text{Log2}(cbWidth) \quad (8-688)$$

$$\log2CbH = \text{Log2}(cbHeight) \quad (8-689)$$

Horizontal change of motion vector dX is derived as follows:

$$10 \quad dX[0] = (cpMvLX[1][0] - cpMvLX[0][0]) \ll (7 - \log2CbW) \quad (8-690)$$

$$dX[1] = (cpMvLX[1][1] - cpMvLX[0][1]) \ll (7 - \log2CbW) \quad (8-691)$$

Vertical change of motion vector dY is derived as follows:

- If $numCpMv$ is equal to 3, dY is derived as follow:

$$dY[0] = (cpMvLX[2][0] - cpMvLX[0][0]) \ll (7 - \log2CbH) \quad (8-692)$$

$$15 \quad dY[1] = (cpMvLX[2][1] - cpMvLX[0][1]) \ll (7 - \log2CbH) \quad (8-693)$$

- Otherwise ($numCpMv$ is equal to 2), dY is derived as follows:

$$dY[0] = -dX[1] \quad (8-694)$$

$$dY[1] = dX[0] \quad (8-695)$$

Motion vector $mvBaseScaled$ corresponding to the top left corner of the luma coding block is derived as follows:

$$20 \quad mvBaseScaled[0] = cpMvLX[0][0] \ll 7 \quad (8-696)$$

$$mvBaseScaled[1] = cpMvLX[0][1] \ll 7 \quad (8-697)$$

1.4.1.2 Step 3. EIF applicability conditions

25 EIF applicability conditions can be for example as following

1. Memory bandwidth restrictions. This kind of restrictions guarantee that the size of area in the reference picture corresponding to the current affine block (EIF block) is no more than predefined threshold T . The example of definition of area in the reference picture corresponding to the current affine block is depicted in FIG. 9.
- 30 2. The internal buffer is limited by R lines, where R is predefined value and can be for example 3, 4, 5 or more. This means that during processing of one line (one row) of

the current block (subblock) no more than R lines from the reference picture can be used.

3. Memory access should be sequential, which means that if for i_{th} line of current block j_{th} line of the reference picture is fetched, then for $(i+1)_{th}$ line of the current block only lines $j+1, j+2, \dots$ can be fetched.
4. No more than one additional line can be fetched for all lines of the current block except for the first.
5. The inequalities for affine motion model parameters, for example as following
 - a.
$$\begin{cases} a \leq dHorX \leq b \\ c \leq dHorY \leq d \\ e \leq dVerX \leq f \\ g \leq dVerY \leq h \end{cases}$$
, where a, b, c, d, e, f, g, h are predefined values or plus/minus infinity.

- b.
$$\begin{cases} dHorX \leq a * dVerX \\ dVerY \leq b * dHorY \end{cases}$$
, where a and b are predefined values

In one example with specific EIF applicability conditions, basic algorithm is as follows:

Algorithm 2. Adaptive usage of EIF and block-based affine motion compensation with specific EIF applicability conditions

1. Calculate optimal subblock size MxN based on affine motion model parameters.
2. If both optimal subblock width M or optimal subblock height N is greater or equal than 8, then perform subblock motion compensation with MxN subblock.
3. Otherwise, check EIF applicability conditions:
 - 3.1. EIF applicability condition 1: The internal buffer is limited by R lines, where R is predefined value and can be for example 3, 4, 5 or more. This means that during processing of one line (one row) of the current block (subblock) no more than R lines from the reference picture can be used.
 - 3.2. EIF applicability condition 2: Memory access should be sequential, which means that if for i_{th} line of current block j_{th} line of the reference picture is fetched, then for $(i+1)_{th}$ line of the current block only lines $j+1, j+2, \dots$ can be fetched.
 - 3.3. EIF applicability condition 3: No more than one additional line can be fetched for all lines of the current block except for the first.
 - 3.4. ...

3.5. EIF applicability condition P

3.6. If all EIF applicability conditions are met, then perform EIF motion compensation.

EIF motion compensation process comprise:

- 3.6.1. Check EIF applicability condition P + 1: Checking memory bandwidth restrictions. This kind of restrictions guarantee that the size of area in the reference picture corresponding to the current affine block (EIF block) is no more than predefined threshold T. The example of definition of area in the reference picture corresponding to the current affine block is depicted in FIG. 9.
- 3.6.2. ...
- 3.6.3. Check EIF applicability condition P+K
- 3.6.4. If one of EIF applicability conditions P+1 .. P+K are not met, then
 - 3.6.4.1. calculate first motion vector range that provide fulfilment of conditions that were not met.
 - 3.6.4.2. set second motion vector range equal to the first motion vector range
- 3.6.5. Otherwise
 - 3.6.5.1. calculate third motion vector range
 - 3.6.5.2. set second motion vector range equal to the third motion vector range
- 3.6.6. Clip motion vectors calculated according to affine model to guarantee, that these vectors are in the second motion vector range.
- 3.7. Otherwise set $M = \max(M, 8)$, $N = \max(N, 8)$ and perform subblock motion compensation with $M \times N$ subblock.

In one exemplary implementation steps 3.1 – 3.5 are implemented as follows:

When predFlagLX is equal to 1, the following applies for X being 0 and 1:

- Horizontal change of motion vector dX, vertical change of motion vector dY and base motion vector mvBaseScaled are derived by invoking the process specified in clause 1.4.1.1.2 with the luma coding block width cbWidth, the luma coding block height cbHeight, number of control point motion vectors numCpMv and the control point motion vectors cpMvLX[cpldx] with cpldx = 0..numCpMv – 1 as inputs.

If dY[1] is less than $(-1) \ll 9$, then the variable eifCanBeAppliedX is equal to FALSE

- Otherwise,
 - If $(\max(0, dY[1]) + \text{Abs}(dX[1])) * (1 + \text{eifSubblockSize})$ is greater than $(1 \ll 9)$ then the variable eifCanBeAppliedX is equal to FALSE.

The variable eifCanBeAppliedX equal to TRUE here means that all EIF applicability conditions 1 – P (steps 3.1 – 3.5 of algorithm 2) are met.

The details about step 3.6.1 of the algorithm 2 are provided below.

5 **1.4.1.2.1 Memory access consumption calculation for affine block in case of usage EIF (step 3.6.1 of algorithm 2)**

The following steps are performed for memory access consumption calculation:

1. Derive the location of each corner sample of the WxH subblock.
2. Derive the location of each corner sample of the subblock used on step 3 of EIF
10 (denote it as EIF intermediate subblock).
3. Derive motion vectors for each corner sample of EIF intermediate subblock.
4. Derive location of transformed subblock in reference picture.
5. Derive the bounding box size for the transformed subblock.
6. Get the memory access consumption based on the bounding box size for transformed
15 subblock size and filter length (EIF uses bilinear interpolation, so filter length is equal to 2).

The details of implementation of these steps are described below.

Step 1. Derive the location of each corner sample of the affine subblock

Denote (x_0, y_0) as the coordinate of the top-left sample of the affine block. In this embodiment for memory access consumption calculation, it's supposed that coordinate of the top-left
20 sample of the affine block is equal to $(1, 1)$. The position (x_0, y_0) does not make sense for memory access consumption calculation and with $(x_0, y_0) = (1, 1)$ the equations are simpler. Then the location of the affine block can be described by the coordinates of its corner samples (top-left, top-right, bottom-left, bottom-right):

$$\begin{cases} (x_0, y_0) = (1, 1) \\ (x_1, y_1) = (W, 1) \\ (x_2, y_2) = (1, H) \\ (x_3, y_3) = (W, H) \end{cases} \quad (3-1)$$

25 **Step 2. Derive the location of each corner sample of the EIF intermediate subblock**

Due to EIF uses 3-tap filter at the step 3, bilinear interpolation at the step 2 of EIF is performed for $(W+2) \times (H+2)$ subblock (one pixel margin from each border is added). This $(W+2) \times (H+2)$ subblock is denoted as intermediate EIF subblock. The coordinates of intermediate EIF block corner samples (top-left, top-right, bottom-left, bottom-right) are:

$$\begin{cases} (X_0, Y_0) = (0, 0) \\ (X_1, Y_1) = (1 + W, 0) \\ (X_2, Y_2) = (0, 1 + H) \\ (X_3, Y_3) = (1 + W, 1 + H) \end{cases} \quad (3-2)$$

The coordinates of corners of affine subblock and of intermediate EIF subblock are depicted in FIG 8.

5 Step 3. Derive motion vectors for each corner sample of EIF intermediate subblock

The initial motion vector (mv_{0x} , mv_{0y}) does not make sense for memory access consumption calculation and with $(mv_{0x}, mv_{0y}) = (dHorX + dVerX, dHorY + dVerY)$ the equations are simpler.

Motion vectors are derived according to equation (1-1).

$$\begin{cases} V_0 = V_{X_0, Y_0} = (0, 0) \\ V_1 = V_{X_1, Y_1} = ((W + 1)dHorX, (W + 1)dHorY) \\ V_2 = V_{X_2, Y_2} = ((H + 1)dVerX, (H + 1)dVerY) \\ V_3 = V_{X_3, Y_3} = ((W + 1)dHorX + (H + 1)dVerX, (W + 1)dHorY + (H + 1)dVerY) \end{cases} \quad (3-3)$$

Step 4. Derive location of transformed block in reference picture

The location of the transformed block in reference picture can be described by the coordinates of its corner samples (top-left, top-right, bottom-left, bottom-right):

$$\begin{cases} (X'_0, Y'_0) = (0, 0) \\ (X'_1, Y'_1) = (1 + W + (W + 1)dHorX, (W + 1)dHorY) \\ (X'_2, Y'_2) = ((H + 1)dVerX, 1 + H + (H + 1)dVerY) \\ (X'_3, Y'_3) = (1 + W + (W + 1)dHorX + (H + 1)dVerX, 1 + H + (W + 1)dHorY + (H + 1)dVerY) \end{cases} \quad (3-4)$$

Step 5. Derive the bounding box size for the transformed subblock

The bounding box size for transformed subblock in reference picture can be calculated by following equation, where max function returns the maximum value of the arguments, min function returns the minimum value of the arguments:

$$\begin{cases} W' = \max(X'_0, X'_1, X'_2, X'_3) - \min(X'_0, X'_1, X'_2, X'_3) + 1 \\ H' = \max(Y'_0, Y'_1, Y'_2, Y'_3) - \min(Y'_0, Y'_1, Y'_2, Y'_3) + 1 \end{cases} \quad (3-5)$$

The location of transformed subblock in reference picture and corresponding bounding box are depicted in FIG. 9.

In one example $W' = \text{Ceil}(W')$, $H' = \text{Ceil}(H')$ is performed after equation (3-5).

In another example $W' = \text{Floor}(W')$, $H' = \text{Floor}(H')$ is performed after equation (3-5).

5 Step 6. Get the memory access consumption

The memory access consumption of the affine subblock in one reference picture can be decided, by the bounding box size for the transformed subblock size and the length of MC interpolation filter for affine motion block T' , e.g., 2, 4, 6, 8, ...:

$$Mem = (W' + T' - 1) * (H' + T' - 1) \quad (3-6)$$

- 10 For EIF, bilinear interpolation is used, therefore filter length is 2 and memory access consumption is equal to

$$Mem = (W' + 1) * (H' + 1) \quad (3-7)$$

1.4.1.2.2 Affine motion model restrictions for EIF (step 3.6.1 of algorithm 2)

- 15 Denote target worst case memory bandwidth as $T = \frac{S_{wc}}{W * H}$, where W and H are current subblock width and heights respectively, and S_{wc} is maximum allowed memory access consumption for the current subblock, according to the target case memory bandwidth. To guarantee that EIF memory bandwidth is not greater than the target case memory bandwidth, the memory access consumption of EIF subblock should be constrained as following condition:

$$\begin{aligned} & \frac{(W' + 1) * (H' + 1)}{W * H} \leq T \\ 20 \quad & \text{or} \\ & (W' + 1) * (H' + 1) \leq T * W * H \quad (3-8) \\ & \text{or} \\ & (W' + 1) * (H' + 1) \leq S_{wc} \end{aligned}$$

- 25 The value T can be predefined in both encoder and decoder, or specified in a parameter set of a codec video sequence, e.g., sequence levels, picture level, slice level parameter set, etc.

In one example, if the maximum allowed memory access consumption per sample is defined as the memory access consumption of a 4x4 block, then T can be derived as follows, where T is the length of interpolation filter:

$$T = \frac{(4+T'-1)*(4+T'-1)}{4*4} \quad (3-9)$$

For T' equal to 6 the restriction is as follows:

$$(W' + 1) * (H' + 1) \leq (4 + 6 - 1) * (4 + 6 - 1) \quad (3-10)$$

In another example, if the maximum allowed memory access consumption per sample is defined as the memory access consumption of a 8x8 block, then T can be derived as follows, where T is the length of interpolation filter:

$$T = \frac{(8+T'-1)*(8+T'-1)}{8*8} \quad (3-11)$$

In another example, the maximum allowed memory access consumption of per sample can be different according the prediction direction of current block, i.e.: when current block is uni-prediction, use threshold T_{UNI} , when current block is bi-prediction, use threshold T_{BI} .

For example, T_{UNI} is defined as the memory access consumption of 4x4 block, T_{BI} is defined as the memory access consumption of 8x4 block, then:

$$T_{UNI} = \frac{(4+T'-1)*(4+T'-1)}{4*4} \quad (3-12)$$

$$T_{BI} = \frac{(8+T'-1)*(4+T'-1)}{8*4} \quad (3-13)$$

In another example, T_{UNI} is defined as the memory access consumption of 4x4 block, T_{BI} is defined as the memory access consumption of 8x8 block, then:

$$T_{UNI} = \frac{(4+T'-1)*(4+T'-1)}{4*4} \quad (3-14)$$

$$T_{BI} = \frac{(8+T'-1)*(8+T'-1)}{8*8} \quad (3-15)$$

T' in the examples above is the length of motion compensation (MC) interpolation filter for translational motion block, e.g., 2, 4, 6, 8....

The value of T , T_{UNI} and T_{BI} may depend on width and height of the current block.

For bi-prediction affine block, the above constraint is applied to both list0 and list1, individually.

In another example memory access consumption is calculated for list0 and for list1 as Mem_0 and Mem_1 and sum of these elements is restricted. For example if the T_{BI} is defined as the memory access consumption of 8x8 block the following restriction is used:

$$Mem_0 + Mem_1 \leq 2 * (8 + T' - 1) * (8 + T' - 1) \quad (3-16)$$

If the condition (3-8) is not met, the motion vector range that guarantee required size of the bounding box is derived and during MV calculation in EIF according to equation (1-1), the

horizontal and vertical parts of motion vectors are clipped according to the derived range. The motion vector range derivation is described in section 1.4.1.2.3.

Step 3.6.1 of the algorithm 2 can be implemented as follows:

The variable clipMVX is derived as follows for X being 0 and 1:

- 5 – clipMVX is set equal to FALSE

The variable ctfSubblockSize is set equal to 4.

When predFlagLX is equal to 1, the following applies for X being 0 and 1:

- Horizontal change of motion vector dX, vertical change of motion vector dY and base motion vector mvBaseScaled are derived by invoking the process specified in clause 1.4.1.1.2 with the luma coding block width cbWidth, the luma coding block height cbHeight, number of control point motion vectors numCpMv and the control point motion vectors cpMvLX[cpIdx] with cpIdx = 0..numCpMv – 1 as inputs.
- 10 – The variable mvWx and mvWy are derived as follows:
- mvWx = max(abs(dX[0]), abs(dX[1]))
- mvWy = max(abs(dY[0]), abs(dY[1]))
- 15 – The variable sizeSbXTemp is specified in Table 8-5 according to the value of mvWx.
- The variable sizeSbYTemp is specified in Table 8-5 according to the value of mvWy.
- The variable sizeSbX is modified as follow:
- sizeSbX = min(sizeSbX, sizeSbXTemp)
- The variable sizeSbY is modified as follow:
- 20 – sizeSbY = min(sizeSbY, sizeSbYTemp)

Table 8-5 – Specification of sizeSbXTemp for various input values of mvWx

mvWx	0	1	2	3	4	> 4
sizeSbX	cbWidth	32	16	8	8	4

Table 8-5 – Specification of sizeSbYTemp for various input values of mvWy

mvWx	0	1	2	3	4	> 4
sizeSbY	cbHeight	32	16	8	8	4

- The variable clipMVX are modified as follows:

- The arrays $X[i]$, $Y[i]$ are derived as follows:
 - $X[0] = 0$
 - $X[1] = (\text{cifSubblockSize} + 1) * (dX[0] + (1 \ll 9))$
 - $X[2] = (\text{cifSubblockSize} + 1) * dY[0]$
 - 5 – $X[3] = X[1] + X[2]$
 - $Y[0] = 0$
 - $Y[1] = (\text{cifSubblockSize} + 1) * dX[1]$
 - $Y[2] = (\text{cifSubblockSize} + 1) * (dY[1] + (1 \ll 9))$
 - $Y[3] = Y[1] + Y[2]$
 - 10 – The variable $X_{\max}$ is set equal to maximum of $X[i]$ for i is equal 0.. 3
 - The variable $X_{\min}$ is set equal to minimum of $X[i]$ for i is equal 0.. 3
 - The variable $Y_{\max}$ is set equal to maximum of $Y[i]$ for i is equal 0.. 3
 - The variable $Y_{\min}$ is set equal to minimum of $Y[i]$ for i is equal 0.. 3
 - The variable W is set equal to $(X_{\max} - X_{\min} + (1 \ll 9) - 1) \gg 9$
 - 15 – The variable H is set equal to $(Y_{\max} - Y_{\min} + (1 \ll 9) - 1) \gg 9$
 - If $(W + 2) * (H + 2)$ is greater than 81, the variable clipMVX is set equal to TRUE

The variables `cifCanBeApplied` and `clipMV` are derived as following:

- $\text{clipMV} = \text{clipMV0} \mid \text{clipMV1}$

- 20 The variable `clipMV` equal to TRUE here means that memory bandwidth EIF applicability condition is not met and further mv clipping is needed (steps 3.6.4 – 3.6.6 of the Algorithm 2)

The section below describes the conventional method of the motion vector range updating near the picture boundaries. The examples of applying the conventional method are depicted in FIG. 10 - 14 at the left part (under the title “Conventional design”).

1.4.1.2.3 Derivation of clipping parameters for affine motion vector (steps 3.6.4 – 3.6.5 of the algorithm 2)

Inputs to this process are:

- a location (xCb, yCb) in full-sample units,
- 5 – two variables cbWidth and cbHeight specifying the width and the height of the current coding block,
- horizontal change of motion vector dX,
- vertical change of motion vector dY,
- motion vector mvBaseScaled,
- width of the picture in samples pic_width,
- 10 – height of the picture in samples pic_height,
- clipMV flag specifying if MV clipping is to be applied (this flag specify whether EIF memory access conditions are fulfilled or not)

Outputs of this process are:

- hor_max, ver_max, hor_min and ver_min that denotes the maximum and minimum allowed motion vector
- 15 horizontal and vertical components.

The center motion vector mv_center is derived as follows:

$$mv_center[0] = (mvBaseScaled[0] + dX[0] * (cbWidth >> 1) + dY[0] * (cbHeight >> 1)) \quad (8-743)$$

$$mv_center[1] = (mvBaseScaled[1] + dX[1] * (cbWidth >> 1) + dY[1] * (cbHeight >> 1)) \quad (8-743)$$

- The rounding process for motion vectors as specified in clause 8.5.3.10 is invoked with mv_center, rightShift set equal to 5, and leftShift set equal to 0 as inputs and the rounded motion vector is return as mv_center.
- 20

The motion vector mv_center is clipped as follows:

$$mv_center[0] = Clip3(-2^{17}, 2^{17} - 1, mv_center[0]) \quad (8-686)$$

$$mv_center[1] = Clip3(-2^{17}, 2^{17} - 1, mv_center[1]) \quad (8-686)$$

The variables hor_max_pic, ver_max_pic, hor_min_pic and ver_min_pic are derived as follows:

$$25 \quad hor_max_pic = (pic_width + 128 - xCb - cbWidth) << 4 \quad (8-743)$$

$$ver_max_pic = (pic_height + 128 - yCb - cbHeight) << 4 \quad (8-743)$$

$$hor_min_pic = (-128 - xCb) << 4 \quad (8-743)$$

$$ver_min_pic = (-128 - yCb) << 4 \quad (8-743)$$

If clipMV is equal to FALSE, then the output variables hor_max, ver_max, hor_min and ver_min that denotes the maximum and minimum allowed motion vector horizontal and vertical components are derived as following (this corresponds to step 3.6.5 of the algorithm 2):

```

hor_max = hor_max_pic << 5 (8-743)
5 ver_max = ver_max_pic << 5 (8-743)
hor_min = hor_min_pic << 5 (8-743)
ver_min = ver_min_pic, << 5 (8-743)

```

Otherwise if clipMV is equal to TRUE, the following steps are applied (this corresponds to step 3.6.4 of the algorithm 2):

```

10 - The variables mv_hor_min, mv_ver_min, mv_hor_max and mv_ver_max are derived as following:
    mv_hor_min = mv_center [ 0 ] - deviationMV [ log2CbWidth - 3 ] (8-743)
    mv_ver_min = mv_center [ 1 ] - deviationMV [ log2CbHeight - 3 ] (8-743)
    mv_hor_max = mv_center [ 0 ] + deviationMV [ log2CbWidth - 3 ] (8-743)
    mv_ver_max = mv_center [ 1 ] + deviationMV [ log2CbHeight - 3 ] (8-743)
15 with array deviationMV specified for k=0..4 as deviationMV[k] = { 64, 128, 272, 560, 1136 }.
    - The variables hor_max, ver_max, hor_min and ver_min are derived as following:
    mv_hor_max = max(mv_hor_max, hor_min) (8-743)
    mv_ver_max = max(mv_ver_max, ver_min) (8-743)
    mv_hor_min = min(mv_hor_min, hor_max) (8-743)
20 mv_ver_min = min(mv_ver_min, ver_max) (8-743)
    hor_max = min(hor_max_pic, mv_hor_max) << 5 (8-743)
    ver_max = min(ver_max_pic, mv_ver_max) << 5 (8-743)
    hor_min = max(hor_min_pic, mv_hor_min) << 5 (8-743)
    ver_min = max(ver_min_pic, mv_ver_min) << 5 (8-743)

```

2 Problems of the conventional design

The conventional design of MV clipping parameters derivation has the following problems:

1. Motion vector range is significantly reduced if central motion vector is close to the picture boundaries
- 5 For example if central motion vector points outside the picture boundary, the motion vector range shrinks significantly as depicted in FIG 10, 11. In some corner cases motion vector range can even collapse to the one point as depicted in FIG 14. Motion vector range reduction cause to motion model degradation. So, instead of the real motion field (“true motion”) (e.g. affine motion like rotation or zooming) prediction signal is generated based on the motion field
- 10 corrupted by clippings to the small motion vector range. This in turn causes to prediction signal quality degradation, which results to bitrate increasing (due to more bits is needed to the residual signal coding) or reconstruction signal quality degradation.

3 Detailed description of the invention

3.1 Motion vector range derivation near the picture boundaries

- 15 In conventional design of MV clipping range derivation for EIF, MV range is significantly reduced in case of central MV, calculated in section 1.4.1.2.3, is close to the picture boundary or outside of the picture boundary. For example, for horizontal direction, if MV spread is greater than the distance from $(x_0 + mv_center[0])$ to the left picture boundary, motion vector range will be reduced as depicted in FIG. 12. If $(x_0 + mv_center[0])$ refer to the position outside
- 20 the picture boundary (e.g. $x_0 + mv_center[0] < 0$ or $x_0 + mv_center[0] < -MAX_CU_SIZE$ in some implementations) the MV range will be reduced to only one point as depicted in FIG. 13. The picture for the right picture boundary is symmetric.

- To avoid reducing MV range in described specific cases the following solution is proposed. In case of MV range, calculated based on central motion vector (mv_center), comprise MVs
- 25 pointing outside the picture boundary (outside the picture boundary with MAX_CU_SIZE margin in some implementations) MV range is calculated based on picture boundary instead of central motion vector. The examples of using new design are depicted in FIG. 12 and FIG. 13. The exemplary implementation of MV range calculation (section 1.4.1.2.3) is described below.

Derivation of clipping parameters for affine motion vector (implementation 1)

Inputs to this process are:

- a location (xCb, yCb) in full-sample units,
- two variables cbWidth and cbHeight specifying the width and the height of the current coding block,
- 5 – horizontal change of motion vector dX,
- vertical change of motion vector dY,
- motion vector mvBaseScaled,
- width of the picture in samples pic_width,
- height of the picture in samples pic_height,
- 10 – clipMV flag specifying MV clipping type.

Outputs of this process are:

- hor_max, ver_max, hor_min and ver_min that denotes the maximum and minimum allowed motion vector horizontal and vertical components.

The variables log2CbW and log2CbH are derived as follows:

$$15 \quad \log2CbWidth = \text{Log2}(cbWidth) \quad (8-774)$$

$$\log2CbHeight = \text{Log2}(cbHeight) \quad (8-775)$$

The variables hor_max_pic, ver_max_pic, hor_min_pic and ver_min_pic are derived as follows:

$$\text{hor_max_pic} = (\text{pic_width} + 128 - \text{xCb} - \text{cbWidth} - 1) \ll 5 \quad (8-776)$$

$$\text{ver_max_pic} = (\text{pic_height} + 128 - \text{yCb} - \text{cbHeight} - 1) \ll 5 \quad (8-777)$$

$$20 \quad \text{hor_min_pic} = (-128 - \text{xCb}) \ll 5 \quad (8-778)$$

$$\text{ver_min_pic} = (-128 - \text{yCb}) \ll 5 \quad (8-779)$$

The variables hor_max_pic, ver_max_pic, hor_min_pic and ver_min_pic represent the second motion vector range that is used to check whether a motion vector points inside the “first area including reference picture”.

The center motion vector mv_center is derived as follows:

$$25 \quad \text{mv_center}[0] = (\text{mvBaseScaled}[0] + \text{dX}[0] * (\text{cbWidth} >> 1) + \text{dY}[0] * (\text{cbHeight} >> 1)) \quad (8-780)$$

$$\text{mv_center}[1] = (\text{mvBaseScaled}[1] + \text{dX}[1] * (\text{cbWidth} >> 1) + \text{dY}[1] * (\text{cbHeight} >> 1)) \quad (8-781)$$

A step of “obtaining a center motion vector, mv_center of a coding block” in the present disclosure corresponds to the equations (8-780) and (8-781).

- 30 The rounding process for motion vectors as specified in clause 8.5.3.10 is invoked with mv_center, rightShift set equal to 4, and leftShift set equal to 0 as inputs and the rounded motion vector is return as mv_center.

If clipMV is equal to FALSE, then the output variables hor_max, ver_max, hor_min and ver_min that denotes the maximum and minimum allowed motion vector horizontal and vertical components are set equal to variables hor_max_pic, ver_max_pic, hor_min_pic and ver_min_pic respectively.

Otherwise, the following steps are applied:

- 35 – The array deviationMV is set equal to { 128, 256, 544, 1120, 2272 }.
- The variables mv_hor_min, mv_ver_min, mv_hor_max and mv_ver_max (hor_min, ver_min, hor_max and ver_max, used in the below) are derived as following:

$$\text{hor_min} = \text{mv_center}[0] - \text{deviationMV}[\log2CbWidth - 3] \quad (8-788)$$

$$\text{ver_min} = \text{mv_center}[1] - \text{deviationMV}[\log2CbHeight - 3] \quad (8-789)$$

$$\text{hor_max} = \text{mv_center}[0] + \text{deviationMV}[\log_2\text{CbWidth} - 3] \quad (8-790)$$

$$\text{ver_max} = \text{mv_center}[1] + \text{deviationMV}[\log_2\text{CbHeight} - 3] \quad (8-791)$$

- 5 A step of “deriving a first motion vector range for the coding block based on the center motion vector and a motion vector spread, wherein the motion vector spread depends on a size of the coding block” in the present disclosure corresponds to the equations (8-788) – (8-791). Motion vector spread is represented by a horizontal motion vector spread $\text{deviationMV}[\log_2\text{CbHeight} - 3]$ and a vertical motion vector spread $[\log_2\text{CbHeight} - 3]$.

– If hor_min is less than hor_min_pic , the variables hor_min and hor_max are updated as following:

$$\text{hor_min} = \text{hor_min_pic} \quad (8-792)$$

$$10 \quad \text{hor_max} = \min(\text{hor_max_pic}, \text{hor_min_pic} + 2 * \text{deviationMV}[\log_2\text{CbWidth} - 3]) \quad (8-793)$$

– Otherwise, if hor_max is greater than hor_max_pic , the variables hor_min and hor_max are updated as following:

$$\text{hor_min} = \max(\text{hor_min_pic}, \text{hor_max_pic} - 2 * \text{deviationMV}[\log_2\text{CbWidth} - 3]) \quad (8-794)$$

$$\text{hor_max} = \text{hor_max_pic} \quad (8-795)$$

- 15 – If ver_min is less than ver_min_pic , the variables ver_min and ver_max are updated as following:

$$\text{ver_min} = \text{ver_min_pic} \quad (8-796)$$

$$\text{ver_max} = \min(\text{ver_max_pic}, \text{ver_min_pic} + 2 * \text{deviationMV}[\log_2\text{CbHeight} - 3]) \quad (8-797)$$

– Otherwise, if ver_max is greater than ver_max_pic , the variables ver_min and ver_max are updated as following:

$$20 \quad \text{ver_min} = \max(\text{ver_min_pic}, \text{ver_max_pic} - 2 * \text{deviationMV}[\log_2\text{CbHeight} - 3]) \quad (8-798)$$

$$\text{ver_max} = \text{ver_max_pic} \quad (8-799)$$

A step of “if the first motion vector range is at least partially pointing outside a first area including a reference picture, updating the first motion vector range to be pointing within the first area” in the present disclosure corresponds to the equations (8-792) – (8-799) and the corresponding conditions above.

- 25 The output variables hor_max , ver_max , hor_min and ver_min are clipped as follows:

$$\text{hor_max} = \text{Clip3}(-2^{17}, 2^{17} - 1, \text{hor_max}) \quad (8-800)$$

$$\text{ver_max} = \text{Clip3}(-2^{17}, 2^{17} - 1, \text{ver_max}) \quad (8-801)$$

$$\text{hor_min} = \text{Clip3}(-2^{17}, 2^{17} - 1, \text{hor_min}) \quad (8-802)$$

$$\text{ver_min} = \text{Clip3}(-2^{17}, 2^{17} - 1, \text{ver_min}) \quad (8-803)$$

- 30 A step of “performing a clipping operation on the updated first motion vector range to be within the range $[-2^{17}, 2^{17}-1]$ ” in the present disclosure corresponds to the equations (8-800) – (8-803).

It can be noted that the second motion vector range is represented by hor_max_pic , ver_max_pic , hor_min_pic and ver_min_pic in equations (8-776) to (8-779).

- 35 It can be noted that the first motion vector range is represented by hor_min , ver_min , hor_max and ver_max in equations (8-788) to (8-791).

It can be noted that the updated first motion vector range is represented by hor_min, ver_min, hor_max and ver_max in equations (8-792) to (8-799).

- 5 It can be noted that the variables hor_min, ver_min, hor_max and ver_max located as right in equations (8-800) to (8-803) representing correspondingly the first minimum MV horizontal component value, the first minimum MV vertical component value, the first maximum MV horizontal component value and the first maximum MV vertical component value of the updated first motion vector range.

- 10 It can be noted that the variables hor_min, ver_min, hor_max and ver_max located as left in equations (8-800) to (8-803) representing correspondingly the first minimum MV horizontal component value, the first minimum MV vertical component value, the first maximum MV horizontal component value and the first maximum MV vertical component value of the updated and clipped first motion vector range.

Derivation of clipping parameters for affine motion vector (implementation 2)

Inputs to this process are:

- 15 – a location (xCb, yCb) in full-sample units,
 – two variables cbWidth and cbHeight specifying the width and the height of the current coding block,
 – horizontal change of motion vector dX,
 – vertical change of motion vector dY,
 – motion vector mvBaseScaled,
 20 – width of the picture in samples pic_width,
 – height of the picture in samples pic_height,
 – clipMV flag specifying MV clipping type.

Outputs of this process are:

- 25 – hor_max, ver_max, hor_min and ver_min that denotes the maximum and minimum allowed motion vector horizontal and vertical components.

The variables log2CbW and log2CbH are derived as follows:

$$\log2CbWidth = \text{Log2}(cbWidth) \quad (8-774)$$

$$\log2CbHeight = \text{Log2}(cbHeight) \quad (8-775)$$

The variables hor_max_pic, ver_max_pic, hor_min_pic and ver_min_pic are derived as follows:

$$30 \quad hor_max_pic = (pic_width + 128 - xCb - cbWidth - 1) << 5 \quad (8-776)$$

$$ver_max_pic = (pic_height + 128 - yCb - cbHeight - 1) << 5 \quad (8-777)$$

$$hor_min_pic = (-128 - xCb) << 5 \quad (8-778)$$

$$ver_min_pic = (-128 - yCb) << 5 \quad (8-779)$$

The center motion vector mv_center is derived as follows:

$$35 \quad mv_center[0] = (mvBaseScaled[0] + dX[0] * (cbWidth >> 1) + dY[0] * (cbHeight >> 1)) \quad (8-780)$$

$$mv_center[1] = (mvBaseScaled[1] + dX[1] * (cbWidth >> 1) + dY[1] * (cbHeight >> 1)) \quad (8-781)$$

The rounding process for motion vectors as specified in clause 8.5.3.10 is invoked with mv_center, rightShift set equal to 4, and leftShift set equal to 0 as inputs and the rounded motion vector is return as mv_center.

If clipMV is equal to FALSE, then the output variables hor_max, ver_max, hor_min and ver_min that denotes the maximum and minimum allowed motion vector horizontal and vertical components are set equal to variables hor_max_pic, ver_max_pic, hor_min_pic and ver_min_pic respectively.

Otherwise, the following steps are applied:

- 5 – The array deviationMV is set equal to { 128, 256, 544, 1120, 2272 }.
- The variables mv_hor_min, mv_ver_min, mv_hor_max and mv_ver_max(hor_min, ver_min, hor_max and ver_max, used as below) are derived as following:

$$\text{hor_min} = \text{mv_center}[0] - \text{deviationMV}[\log_2\text{CbWidth} - 3] \quad (8-788)$$

$$\text{ver_min} = \text{mv_center}[1] - \text{deviationMV}[\log_2\text{CbHeight} - 3] \quad (8-789)$$
- 10 – $\text{hor_max} = \text{mv_center}[0] + \text{deviationMV}[\log_2\text{CbWidth} - 3]$ (8-790)

$$\text{ver_max} = \text{mv_center}[1] + \text{deviationMV}[\log_2\text{CbHeight} - 3] \quad (8-791)$$
- If hor_min is less than hor_min_pic, the variables hor_min and hor_max are updated as following:

$$\text{hor_min} = \text{hor_min_pic} \quad (8-788)$$

$$\text{hor_max} = \min(\text{hor_max_pic}, \text{hor_min_pic} + 2 * \text{deviationMV}[\log_2\text{CbWidth} - 3]) \quad (8-790)$$
- 15 – Otherwise, if hor_max is greater than hor_max_pic, the variables hor_min and hor_max are updated as following:

$$\text{hor_min} = \max(\text{hor_min_pic}, \text{hor_max_pic} - 2 * \text{deviationMV}[\log_2\text{CbWidth} - 3]) \quad (8-788)$$

$$\text{hor_max} = \text{hor_max_pic} \quad (8-790)$$
- If ver_min is less than ver_min_pic, the variables ver_min and ver_max are updated as following:

$$\text{ver_min} = \text{ver_min_pic} \quad (8-788)$$

$$\text{ver_max} = \min(\text{ver_max_pic}, \text{ver_min_pic} + 2 * \text{deviationMV}[\log_2\text{CbHeight} - 3]) \quad (8-790)$$
- Otherwise, if ver_max is greater than ver_max_pic, the variables ver_min and ver_max are updated as following:

$$\text{ver_min} = \max(\text{ver_min_pic}, \text{ver_max_pic} - 2 * \text{deviationMV}[\log_2\text{CbHeight} - 3]) \quad (8-788)$$
- 25 – $\text{ver_max} = \text{ver_max_pic}$ (8-790)

The output variables hor_max, ver_max, hor_min and ver_min are clipped as follows:

- $$\text{hor_max} = \text{Clip3}(-2^{17}, 2^{17} - 1, \text{hor_max}) \quad (8-784)$$
- $$\text{ver_max} = \text{Clip3}(-2^{17}, 2^{17} - 1, \text{ver_max}) \quad (8-785)$$
- $$\text{hor_min} = \text{Clip3}(-2^{17}, 2^{17} - 1, \text{hor_min}) \quad (8-786)$$
- 30 – $\text{ver_min} = \text{Clip3}(-2^{17}, 2^{17} - 1, \text{ver_min}) \quad (8-787)$

The difference between implementation 1 and implementation 2 is in picture boundaries derivation. In implementation 1 it's supposed, that there is a margin of 128 (MAX_CU_SIZE) pixels around the picture. The margin size can depend on coding tree unit size.

3.2 Boundary motion vector clipping

- 5 In EVC it's supposed that each of MV component can be stored using 18 bits. To fulfil this general requirement the current invention adds clipping to the range $[-2^{17}, 2^{17} - 1]$ as a final step of boundary motion vector derivation. To perform this clipping without losing the motion vector magnitude, operations in section 1.4.1.2.3 are performed in 5-bit precision instead of 9-bit precision. With these changes the following changes in general EIF algorithm (section 1.3)
- 10 are needed (newly added parts are marked ~~in grey~~, removed parts are marked as ~~strikethrough~~)

– The motion vector mvX is derived as follows :

$$\text{mvX}[0] = (\text{mvBaseScaled}[0] + \text{dX}[0] * x + \text{dY}[0] * y) \gg 4 \quad (8-758)$$

$$\text{mvX}[1] = (\text{mvBaseScaled}[1] + \text{dX}[1] * x + \text{dY}[1] * y) \gg 4 \quad (8-759)$$

$$\text{mvX}[0] = \text{Clip3}(\text{hor_min}, \text{hor_max}, \text{mvX}[0]) \quad (8-760)$$

$$\text{mvX}[1] = \text{Clip3}(\text{ver_min}, \text{ver_max}, \text{mvX}[1]) \quad (8-761)$$

A step of “performing a clipping operation on a motion vector of a pixel of the coding block to be within a range, to obtain a clipped motion vector, wherein the range depends on the updated first motion vector range” in the present disclosure corresponds to the equations (8-760) to (8-761).

– The variables xInt, yInt, xFrac and yFrac are derived as follows:

$$20 \quad \text{xInt} = \text{xCb} + (\text{mvX}[0] \gg 9) + x \quad (8-762)$$

$$\text{yInt} = \text{yCb} + (\text{mvX}[1] \gg 9) + y \quad (8-763)$$

$$\text{xFrac} = (\text{mvX}[0] \gg 4) \& 31 \quad (8-764)$$

$$\text{yFrac} = (\text{mvX}[1] \gg 4) \& 31 \quad (8-765)$$

3.3 Memory bandwidth restrictions alignment

- 25 The specification text examples described above use threshold 81 in the implementation of step 6 in section 1.4.1.2.2 for 4x4 EIF subblocks. The section below explains the reasons for using this value of the threshold.

According to motion vector clipping spread, defined by array deviationMV (section 1.4.1.2.3), the memory bandwidth threshold for uni-predictive 8x8 (8x8 is a worst case memory

- 30 bandwidth, according to array deviationMV) affine blocks is equal to $\frac{(4+4+8+1)^2}{8*8} \approx 4.5$, that corresponds value 9 for the bi-predictive blocks.

The threshold 81, used in the implementation of step 6 in section 1.4.1.2.2 for 4x4 EIF subblocks, corresponds to memory bandwidth $\frac{81}{4 \times 4} \approx 5$ for uni-predictive block that corresponds to the value 10 for bi-predictive blocks.

To align the thresholds used in sections 1.4.1.2.2 and 1.4.1.2.3 the current invention change the threshold in section 1.4.1.2.2 from 81 to 72 to achieve memory bandwidth threshold equal to 4.5 for uni-predictive and 9 for bi-predictive blocks. With such change memory bandwidth restrictions in sections 1.4.1.2.2 and 1.4.1.2.3 are aligned.

4 The examples of specification text of EVC with the present invention

8.5.3.8 Derivation process for affine subblock size

Inputs to this process are:

- two variables cbWidth and cbHeight specifying the width and the height of the luma coding block,
- the number of control point motion vectors numCpMv,
- the control point motion vectors cpMvLX[cpIdx], with cpIdx = 0..numCpMv – 1 and X being 0 or 1,
- the prediction list utilization flags predFlagLX, with X being 0 or 1.

Outputs of this process are:

- the size of luma coding subblocks in horizontal direction sizeSbX and in vertical direction sizeSbY,
- the number of luma coding subblocks in horizontal direction numSbX and in vertical direction numSbY,
- the flag clipMV indicating motion vector clipping type for blocks processed with EIF.

sizeSbX is set equal to cbWidth, sizeSbY is set equal to cbHeight.

The variables eifCanBeAppliedX and clipMVX are derived as follows for X being 0 and 1:

- eifCanBeAppliedX is set to TRUE
- clipMVX is set equal to FALSE

The variable eifSubblockSize is set equal to 4.

When predFlagLX is equal to 1, the following applies for X being 0 and 1:

- Horizontal change of motion vector dX, vertical change of motion vector dY and base motion vector mvBaseScaled are derived by invoking the process specified in clause 8.5.3.9 with the luma coding block width cbWidth, the luma coding block height cbHeight, number of control point motion vectors numCpMv and the control point motion vectors cpMvLX[cpIdx] with cpIdx = 0..numCpMv – 1 as inputs.
- The variable mvWx and mvWy are derived as follows:
 - $mvWx = \max(\text{abs}(dX[0]), \text{abs}(dX[1]))$
 - $mvWy = \max(\text{abs}(dY[0]), \text{abs}(dY[1]))$
- The variable sizeSbXTemp is specified in Table 8-10 according to the value of mvWx.
- The variable sizeSbYTemp is specified in Table 8-11 according to the value of mvWy.

- The variable sizeSbX is modified as follow:
 - $\text{sizeSbX} = \min(\text{sizeSbX}, \text{sizeSbXTemp})$
- The variable sizeSbY is modified as follow:
 - $\text{sizeSbY} = \min(\text{sizeSbY}, \text{sizeSbYTemp})$

5

Table 8-10 – Specification of sizeSbXTemp for various input values of mvWx

mvWx	0	1	2	3	4	> 4
sizeSbX	cbWidth	32	16	8	8	4

Table 8-11 – Specification of sizeSbYTemp for various input values of mvWy

mvWy	0	1	2	3	4	> 4
sizeSbY	cbHeight	32	16	8	8	4

- The variables eifCanBeAppliedX and clipMVX are modified as follows:
 - The arrays X[i], Y[i] are derived as follows:
 - 10 – $X[0] = 0$
 - $X[1] = (\text{eifSubblockSize} + 1) * (dX[0] + (1 \ll 9))$
 - $X[2] = (\text{eifSubblockSize} + 1) * dY[0]$
 - $X[3] = X[1] + X[2]$
 - $Y[0] = 0$
 - 15 – $Y[1] = (\text{eifSubblockSize} + 1) * dX[1]$
 - $Y[2] = (\text{eifSubblockSize} + 1) * (dY[1] + (1 \ll 9))$
 - $Y[3] = Y[1] + Y[2]$
 - The variable Xmax is set equal to maximum of X[i] for i is equal 0.. 3
 - The variable Xmin is set equal to minimum of X[i] for i is equal 0.. 3
 - 20 – The variable Ymax is set equal to maximum of Y[i] for i is equal 0.. 3
 - The variable Ymin is set equal to minimum of Y[i] for i is equal 0.. 3
 - The variable W is set equal to $(Xmax - Xmin + (1 \ll 9) - 1) \gg 9$
 - The variable H is set equal to $(Ymax - Ymin + (1 \ll 9) - 1) \gg 9$
 - If $(W + 2) * (H + 2)$ is greater than 32, the variable clipMVX is set equal to TRUE
 - 25 – If $dY[1]$ is less than $(-1 \ll 9)$, then the variable eifCanBeAppliedX is equal to FALSE
 - Otherwise,
 - If $(\max(0, dY[1]) + \text{Abs}(dX[1])) * (1 + \text{eifSubblockSize})$ is greater than $(1 \ll 9)$ then the variable eifCanBeAppliedX is equal to FALSE.

The variables `EIFCanBeApplied` and `clipMV` are derived as following:

- `EIFCanBeApplied` = `EIFCanBeApplied0` & `EIFCanBeApplied1`
- `clipMV` = `clipMV0` | `clipMV1`

If `EIFCanBeApplied` is equal to `FALSE` than the variables `sizeSbX` and `sizeSbY` are modified as follows:

- 5 – `sizeSbX` = $\max(8, \text{sizeSbX})$
- `sizeSbY` = $\max(8, \text{sizeSbY})$

The number of luma coding subblocks in horizontal direction `numSbX` and in vertical direction `numSbY` are derived as follows:

- `numSbX` = `cbWidth` / `sizeSbX`
- 10 – `numSbY` = `cbHeight` / `sizeSbY`
- [...]

8.5.4.3 Interpolation process for the enhanced interpolation filter

Inputs to this process are:

- a location (`xCb`, `yCb`) in full-sample units,
- 15 – two variables `cbWidth` and `cbHeight` specifying the width and the height of the current coding block,
- horizontal change of motion vector `dX`,
- vertical change of motion vector `dY`,
- motion vector `mvBaseScaled`,
- the selected reference picture sample arrays `refPicLX`,
- 20 – sample bit depth `bitDepth`
- width of the picture in samples `pic_width`,
- height of the picture in samples `pic_height`,
- `clipMV` flag specifying MV clipping type,
- `isLuma` flag specifying whether luma or chroma is processed..

25 Outputs of this process are:

- an (`cbWidth` / `SubWidthC`)x(`cbHeight` / `SubHeightC`) array `predSamplesLX` of prediction sample values.

Interpolation filter coefficients $T[p]$ for each fractional sample position p equal to `xFrac` or `yFrac` are specified in Table 8-16.

- 30 The variables `hor_max`, `ver_max`, `hor_min` and `ver_min` are derived by invoking the process specified in 8.5.4.5 with a location (`xCb`, `yCb`) in full-sample units, two variables `cbWidth` and `cbHeight` specifying the width and the height of the current coding block, horizontal change of motion vector `dX`, vertical change of motion vector `dY`, motion vector `mvBaseScaled`, width of the picture in samples `pic_width`, height of the picture in samples `pic_height` and `clipMV` flag as input, and `hor_max`, `ver_max`, `hor_min` and `ver_min` as output.

- 35 If `isLuma` is equal to `FALSE`, then variables `mvBaseScaled`, `hor_min`, `hor_max`, `ver_min`, `ver_max` are modified as follows:

- $\text{xCb} = \text{xCb} / \text{SubWidthC}$ (8-748)
- $\text{yCb} = \text{yCb} / \text{SubHeightC}$ (8-749)
- $\text{cbWidth} = \text{cbWidth} / \text{SubWidthC}$ (8-750)

- $cbHeight = cbHeight / SubHeightC$ (8-751)
- $mvBaseScaled[0] = mvBaseScaled[0] / SubWidthC$ (8-752)
- $mvBaseScaled[1] = mvBaseScaled[1] / SubHeightC$ (8-753)
- $hor_min = hor_min / SubWidthC$ (8-754)
- 5 – $hor_max = hor_max / SubWidthC$ (8-755)
- $ver_min = ver_min / SubHeightC$ (8-756)
- $ver_max = ver_max / SubHeightC$ (8-757)

10 The variables hor_min , hor_max , ver_min , ver_max before applying equations (8-754 – 8-757) represents the updated first motion vector range. The variables hor_min , hor_max , ver_min , ver_max after applying equations (8-754 – 8-757) represent the range that will be used in a step “performing a clipping operation on a motion vector of a pixel of the coding block to be within a range, to obtain a clipped motion vector, wherein the range depends on the updated first motion vector range”.

The variables $shift0$, $shift1$, $offset0$ and $offset1$ are derived as follows:

$shift0$ is set equal to $bitDepth - 8$, $offset0$ is equal to 0,

- 15 $shift1$ is set equal to $12 - shift0$, $offset1$ is equal to $2^{shift1-1}$

For $x = -1..cbWidth$ and $y = -1..cbHeight$, the following applies:

- The motion vector mvX is derived as follows :
 - $mvX[0] = (mvBaseScaled[0] + dX[0] * x + dY[0] * y) >> 4$ (8-758)
 - $mvX[1] = (mvBaseScaled[1] + dX[1] * x + dY[1] * y) >> 4$ (8-759)
 - 20 $mvX[0] = Clip3(hor_min, hor_max, mvX[0])$ (8-760)
 - $mvX[1] = Clip3(ver_min, ver_max, mvX[1])$ (8-761)

A step of “performing a clipping operation on a motion vector of a pixel of the coding block to be within a range, to obtain a clipped motion vector, wherein the range depends on the updated first motion vector range” in the present disclosure corresponds to the equations (8-760) to (8-761).

25

The equations below corresponds to the step of “performing pixel-based motion compensation based on the clipped motion vector”.

- The variables $xInt$, $yInt$, $xFrac$ and $yFrac$ are derived as follows:
 - $xInt = xCb + (mvX[0] >> 5) + x$ (8-762)
 - 30 $yInt = yCb + (mvX[1] >> 5) + y$ (8-763)
 - $xFrac = mvX[0] \& 31$ (8-764)
 - $yFrac = mvX[1] \& 31$ (8-765)
- The variables A and B are derived as follows:
 - $A = (refPicLX[xInt][yInt] * T[xFrac][0] +$
 - 35 $+ refPicLX[xInt + 1][yInt] * T[xFrac][1] + offset0) >> shift0$ (8-766)
 - $B = (refPicLX[xInt][yInt + 1] * T[xFrac][0] +$
 - $+ refPicLX[xInt + 1][yInt + 1] * T[xFrac][1] + offset0) >> shift0$ (8-767)
- The sample value $b_{x,y}$ corresponding to location (x, y) is derived as follows:

$$b_{x,y} = (A * T[yFrac][0] + B * T[yFrac][1] + offset1) \gg shift1 \quad (8-768)$$

The enhancement interpolation filter coefficients $eF[]$ are specified as $\{-1, 10, -1\}$.

The variables $shift2$, $shift3$, $offset2$ and $offset3$ are derived as follows:

- 5 $shift2$ is set equal to $\max(bit_depth - 11, 0)$, $offset2$ is equal to $2^{shift2-1}$,
 $shift3$ is set equal to $(6 - \max(bit_depth - 11, 0))$, $offset3$ is equal to $2^{shift3-1}$,

For $x = 0..cbWidth - 1$ and $y = -1..cbHeight$, the following applies:

$$h_{x,y} = (eF[0] * b_{x-1,y} + eF[1] * b_{x,y} + eF[2] * b_{x+1,y} + offset2) \gg shift2 \quad (8-769)$$

For $x = 0..cbWidth - 1$ and $y = 0..cbHeight - 1$, the following applies:

$$10 \quad \begin{aligned} & \text{predSamplesLXL}[x][y] = \text{Clip3}(0, (1 \ll bitDepth) - 1, \\ & (eF[0] * h_{x,y-1} + eF[1] * h_{x,y} + eF[2] * h_{x,y+1} + offset3) \gg shift3) \end{aligned} \quad (8-770)$$

Table 8-16 – Specification of the interpolation filter coefficients $T[p]$ for each fractional sample position p

Fractional sample position p	Interpolation filter coefficients	
	$T[p][0]$	$T[p][1]$
0	64	0
1	62	2
2	60	4
3	58	6
4	56	8
5	54	10
6	52	12
7	50	14
8	48	16
9	46	18

10	44	20
11	42	22
12	40	24
13	38	26
14	36	28
15	34	30
16	32	32
17	30	34
18	28	36
19	26	38
20	24	40
21	22	42
22	20	44
23	18	46
24	16	48
25	14	50
26	12	52
27	10	54
28	8	56
29	6	58

30	4	60
31	2	62

8.5.4.5 Derivation of clipping parameters for affine motion vector (variant 1)

Inputs to this process are:

- 5 – a location (xCb, yCb) in full-sample units,
- two variables cbWidth and cbHeight specifying the width and the height of the current coding block,
- horizontal change of motion vector dX,
- vertical change of motion vector dY,
- motion vector mvBaseScaled,
- 10 – width of the picture in samples pic_width,
- height of the picture in samples pic_height,
- clipMV flag specifying MV clipping type.

Outputs of this process are:

- 15 – hor_max, ver_max, hor_min and ver_min that denotes the maximum and minimum allowed motion vector horizontal and vertical components.

The variables log2CbW and log2CbH are derived as follows:

$$\log2CbWidth = \text{Log2}(cbWidth) \quad (8-774)$$

$$\log2CbHeight = \text{Log2}(cbHeight) \quad (8-775)$$

The variables hor_max_pic, ver_max_pic, hor_min_pic and ver_min_pic are derived as follows:

$$20 \quad \text{hor_max_pic} = (\text{pic_width} + 128 - \text{xCb} - \text{cbWidth} - 1) \ll 5 \quad (8-776)$$

$$\text{ver_max_pic} = (\text{pic_height} + 128 - \text{yCb} - \text{cbHeight} - 1) \ll 5 \quad (8-777)$$

$$\text{hor_min_pic} = (-128 - \text{xCb}) \ll 5 \quad (8-778)$$

$$\text{ver_min_pic} = (-128 - \text{yCb}) \ll 5 \quad (8-779)$$

- 25 The variables hor_max_pic, ver_max_pic, hor_min_pic and ver_min_pic represent the second motion vector range that is used to check whether a motion vector points inside the “first area including reference picture”.

The center motion vector mv_center is derived as follows:

$$\text{mv_center}[0] = (\text{mvBaseScaled}[0] + \text{dX}[0] * (\text{cbWidth} \gg 1) + \text{dY}[0] * (\text{cbHeight} \gg 1)) \quad (8-780)$$

$$\text{mv_center}[1] = (\text{mvBaseScaled}[1] + \text{dX}[1] * (\text{cbWidth} \gg 1) + \text{dY}[1] * (\text{cbHeight} \gg 1)) \quad (8-781)$$

- 30 A step of “obtaining a center motion vector, mv_center of a coding block” in the present disclosure corresponds to the equations (8-780) and (8-781).

The rounding process for motion vectors as specified in clause 8.5.3.10 is invoked with mv_center, rightShift set equal to 4, and leftShift set equal to 0 as inputs and the rounded motion vector is return as mv_center.

If clipMV is equal to FALSE, then the output variables hor_max, ver_max, hor_min and ver_min that denotes the maximum and minimum allowed motion vector horizontal and vertical components are set equal to variables hor_max_pic, ver_max_pic, hor_min_pic and ver_min_pic respectively.

Otherwise, the following steps are applied:

- 5 – The array deviationMV is set equal to { 128, 256, 544, 1120, 2272 }.
- The variables mv_hor_min, mv_ver_min, mv_hor_max and mv_ver_max (hor_min, ver_min, hor_max and ver_max, used as below) are derived as following:

$$\text{hor_min} = \text{mv_center} [0] - \text{deviationMV} [\log_2 \text{CbWidth} - 3] \quad (8-788)$$

$$\text{ver_min} = \text{mv_center} [1] - \text{deviationMV} [\log_2 \text{CbHeight} - 3] \quad (8-789)$$

10 $\text{hor_max} = \text{mv_center} [0] + \text{deviationMV} [\log_2 \text{CbWidth} - 3] \quad (8-790)$

$$\text{ver_max} = \text{mv_center} [1] + \text{deviationMV} [\log_2 \text{CbHeight} - 3] \quad (8-791)$$

A step of “deriving a first motion vector range for the coding block based on the center motion vector and a motion vector spread, wherein the motion vector spread depends on a size of the coding block” in the present disclosure corresponds to the equations (8-788) – (8-791). Motion vector spread is represented by a horizontal motion vector spread deviationMV[log2CbHeight – 3] and a vertical motion vector spread [log2CbWidth – 3].

- 15 – If hor_min is less than hor_min_pic, the variables hor_min and hor_max are updated as following:

$$\text{hor_min} = \text{hor_min_pic} \quad (8-792)$$

$$\text{hor_max} = \min (\text{hor_max_pic}, \text{hor_min_pic} + 2 * \text{deviationMV} [\log_2 \text{CbWidth} - 3]) \quad (8-793)$$

- 20 – Otherwise, if hor_max is greater than hor_max_pic, the variables hor_min and hor_max are updated as following:

$$\text{hor_min} = \max (\text{hor_min_pic}, \text{hor_max_pic} - 2 * \text{deviationMV} [\log_2 \text{CbWidth} - 3]) \quad (8-794)$$

$$\text{hor_max} = \text{hor_max_pic} \quad (8-795)$$

- If ver_min is less than ver_min_pic, the variables ver_min and ver_max are updated as following:

25 $\text{ver_min} = \text{ver_min_pic} \quad (8-796)$

$$\text{ver_max} = \min (\text{ver_max_pic}, \text{ver_min_pic} + 2 * \text{deviationMV} [\log_2 \text{CbHeight} - 3]) \quad (8-797)$$

- Otherwise, if ver_max is greater than ver_max_pic, the variables ver_min and ver_max are updated as following:

$$\text{ver_min} = \max (\text{ver_min_pic}, \text{ver_max_pic} - 2 * \text{deviationMV} [\log_2 \text{CbHeight} - 3]) \quad (8-798)$$

30 $\text{ver_max} = \text{ver_max_pic} \quad (8-799)$

A step of “if the first motion vector range is at least partially pointing outside a first area including a reference picture, updating the first motion vector range to be pointing within the first area” in the present disclosure corresponds to the equations (8-792) – (8-799) and the corresponding conditions above.

The output variables hor_max, ver_max, hor_min and ver_min are clipped as follows:

35 $\text{hor_max} = \text{Clip3}(-2^{17}, 2^{17} - 1, \text{hor_max}) \quad (8-800)$

$$\text{ver_max} = \text{Clip3}(-2^{17}, 2^{17} - 1, \text{ver_max}) \quad (8-801)$$

$$\text{hor_min} = \text{Clip3}(-2^{17}, 2^{17} - 1, \text{hor_min}) \quad (8-802)$$

$$\text{ver_min} = \text{Clip3}(-2^{17}, 2^{17} - 1, \text{ver_min}) \quad (8-803)$$

A step of “performing a clipping operation on the updated first motion vector range to be within the range $[-2^{17}, 2^{17}-1]$ ” in the present disclosure corresponds to the equations (8-800) – (8-803).

5 8.5.4.5 Derivation of clipping parameters for affine motion vector (variant 2)

Inputs to this process are:

- a location (xCb, yCb) in full-sample units,
- two variables cbWidth and cbHeight specifying the width and the height of the current coding block,
- horizontal change of motion vector dX,
- 10 – vertical change of motion vector dY,
- motion vector mvBaseScaled,
- width of the picture in samples pic_width,
- height of the picture in samples pic_height,
- clipMV flag specifying MV clipping type.
- 15 Outputs of this process are:
- hor_max, ver_max, hor_min and ver_min that denotes the maximum and minimum allowed motion vector horizontal and vertical components.

The variables log2CbW and log2CbH are derived as follows:

$$\begin{aligned} \log2CbWidth &= \text{Log2}(cbWidth) & (8-774) \\ \log2CbHeight &= \text{Log2}(cbHeight) & (8-775) \end{aligned}$$

The variables hor_max_pic, ver_max_pic, hor_min_pic and ver_min_pic are derived as follows:

$$\begin{aligned} \text{hor_max_pic} &= (\text{pic_width} - \text{xCb} - \text{cbWidth} - 1) \ll 5 & (8-776) \\ \text{ver_max_pic} &= (\text{pic_height} - \text{yCb} - \text{cbHeight} - 1) \ll 5 & (8-777) \\ \text{hor_min_pic} &= (- \text{xCb}) \ll 5 & (8-778) \\ 25 \quad \text{ver_min_pic} &= (- \text{yCb}) \ll 5 & (8-779) \end{aligned}$$

The center motion vector mv_center is derived as follows:

$$\text{mv_center}[0] = (\text{mvBaseScaled}[0] + \text{dX}[0] * (\text{cbWidth} \gg 1) + \text{dY}[0] * (\text{cbHeight} \gg 1)) \quad (8-780)$$

$$\text{mv_center}[1] = (\text{mvBaseScaled}[1] + \text{dX}[1] * (\text{cbWidth} \gg 1) + \text{dY}[1] * (\text{cbHeight} \gg 1)) \quad (8-781)$$

30 The rounding process for motion vectors as specified in clause 8.5.3.10 is invoked with mv_center, rightShift set equal to 4, and leftShift set equal to 0 as inputs and the rounded motion vector is return as mv_center.

If clipMV is equal to FALSE, then the output variables hor_max, ver_max, hor_min and ver_min that denotes the maximum and minimum allowed motion vector horizontal and vertical components are set equal to variables hor_max_pic, ver_max_pic, hor_min_pic and ver_min_pic respectively.

Otherwise, the following steps are applied:

- 35 – The array deviationMV is set equal to { 128, 256, 544, 1120, 2272 }.
 - The variables mv_hor_min, mv_ver_min, mv_hor_max and mv_ver_max are derived as following:
- $$\text{hor_min} = \text{mv_center}[0] - \text{deviationMV}[\log2CbWidth - 3] \quad (8-788)$$

$$\text{ver_min} = \text{mv_center} [1] - \text{deviationMV} [\log_2 \text{CbHeight} - 3] \quad (8-789)$$

$$\text{hor_max} = \text{mv_center} [0] + \text{deviationMV} [\log_2 \text{CbWidth} - 3] \quad (8-790)$$

$$\text{ver_max} = \text{mv_center} [1] + \text{deviationMV} [\log_2 \text{CbHeight} - 3] \quad (8-791)$$

– If hor_min is less than hor_min_pic, the variables hor_min and hor_max are updated as following:

$$5 \quad \text{hor_min} = \text{hor_min_pic} \quad (8-788)$$

$$\text{hor_max} = \min (\text{hor_max_pic}, \text{hor_min_pic} + 2 * \text{deviationMV} [\log_2 \text{CbWidth} - 3]) \quad (8-790)$$

– Otherwise, if hor_max is greater than hor_max_pic, the variables hor_min and hor_max are updated as following:

$$\text{hor_min} = \max (\text{hor_min_pic}, \text{hor_max_pic} - 2 * \text{deviationMV} [\log_2 \text{CbWidth} - 3]) \quad (8-788)$$

$$10 \quad \text{hor_max} = \text{hor_max_pic} \quad (8-790)$$

– If ver_min is less than ver_min_pic, the variables ver_min and ver_max are updated as following:

$$\text{ver_min} = \text{ver_min_pic} \quad (8-788)$$

$$\text{ver_max} = \min (\text{ver_max_pic}, \text{ver_min_pic} + 2 * \text{deviationMV} [\log_2 \text{CbHeight} - 3]) \quad (8-790)$$

– Otherwise, if ver_max is greater than ver_max_pic, the variables ver_min and ver_max are updated as following:

$$15 \quad \text{ver_min} = \max (\text{ver_min_pic}, \text{ver_max_pic} - 2 * \text{deviationMV} [\log_2 \text{CbHeight} - 3]) \quad (8-788)$$

$$\text{ver_max} = \text{ver_max_pic} \quad (8-790)$$

The output variables hor_max, ver_max, hor_min and ver_min are clipped as follows:

$$\text{hor_max} = \text{Clip3}(-2^{17}, 2^{17} - 1, \text{hor_max}) \quad (8-784)$$

$$20 \quad \text{ver_max} = \text{Clip3}(-2^{17}, 2^{17} - 1, \text{ver_max}) \quad (8-785)$$

$$\text{hor_min} = \text{Clip3}(-2^{17}, 2^{17} - 1, \text{hor_min}) \quad (8-786)$$

$$\text{ver_min} = \text{Clip3}(-2^{17}, 2^{17} - 1, \text{ver_min}) \quad (8-787)$$

25 FIG. 17 is a block diagram illustrating the method according to the present disclosure. The method for coding video data using an affine motion model comprises the steps of

1701: obtaining a center motion vector of a coding block,

1702: deriving a first motion vector range for the coding block based on the center motion vector and a motion vector spread, wherein the motion vector spread depends on a size of the

30 coding block,

1703: if the first motion vector range is at least partially pointing outside a first area including a reference picture, updating the first motion vector range to be pointing within the first area,

such that a minimum value and/or a maximum value of the updated first motion vector range is pointing at the boundary of the first area, wherein the difference between the maximum value and the minimum value of the updated first motion vector range is equal to the minimum of a double value of the motion vector spread and the size of the first area , and

- 5 1704: performing pixel-based motion compensation based on the updated first motion vector range.

Examples of the condition that the motion vector range is at least partially located outside a boundary of a reference picture are shown in Fig. 12 and 13, in which the motion vector range (spread to the left and right of the central motion vector) is located partially outside the left
 10 boundary of the picture. Similar examples apply to the right, lower and upper boundaries. The method avoids a significant reduction of the motion vector range near the picture boundaries. The motion vector range is here calculated based on a motion vector spread and the reference picture boundary, instead of based on a motion vector spread the central motion vector as in the prior art.

15

The invention provides the following further embodiments:

1. A method for performing enhanced interpolation filtering on a boundary of a picture, comprising:
 - 20 obtaining a motion vector, wherein a target block indicated by the motion vector is a block adjacent to the boundary of the picture;
 - determining a motion vector range based on the motion vector, wherein the motion vector range is determined by the size of the target block, and wherein the motion vector range is inside the picture and starts from the boundary of the picture; and
 - 25 performing enhanced interpolation filtering based on the motion vector range.
2. The method of embodiment 1, wherein the picture is a reference picture of a current block, and wherein the motion vector is a motion vector of the current block.
3. The method of embodiment 1, wherein the picture comprises a margin, wherein the margin is an extended area surrounding an original picture.
- 30 4. The method of embodiment 3, wherein a size of the margin depends on a maximum CTU size.
5. The method of embodiment 1, wherein the motion vector range is calculated based on predefined integer numbers for each preset block width and for each preset block height.

6. A codec comprising processing circuitry for carrying out the method according to any one of embodiments 1 to 5.
7. The method of embodiment 6, wherein the codec comprises an encoder or a decoder.
8. An apparatus for performing enhanced interpolation filtering on a boundary of a picture, comprising:
 - a memory and a processor coupled to the memory, wherein the processor is configured to: obtain a motion vector, wherein a target block indicated by the motion vector is a block adjacent to the boundary of the picture;
 - determine a motion vector range based on the motion vector, wherein the motion vector range is determined by the size of the target block, and wherein the motion vector range is inside the picture and starts from the boundary of the picture; and
 - perform enhanced interpolation filtering based on the motion vector range.
9. The apparatus of embodiment 8, wherein the picture is a reference picture of a current block, and wherein the motion vector is a motion vector of the current block.
10. The apparatus of embodiment 8, wherein the picture comprises a margin, wherein the margin is an extended area surrounding an original picture.
11. The apparatus of embodiment 10, wherein a size of the margin depends on a maximum CTU size.
12. The apparatus of embodiment 8, wherein the motion vector range is calculated based on predefined integer numbers for each preset block width and for each preset block height.

- Embodiment 13. A method of coding implemented by a decoding/encoding device for coding video data, comprising:
- deriving a first motion vector for central point of a affine block,
 - deriving a motion vector spread based on the affine block size (e.g. the affine block size WxH),
 - deriving a first motion vector range based on the first motion vector and the motion vector spread,
 - deriving a second motion vector range in order to guarantee that MVs points to a first predefined area,
 - checking whether the first motion vector range is included into the second motion vector range,
 - if the first motion vector range is not included into the second motion vector range,

- deriving a third motion vector range based on the second motion vector range and the motion vector spread,
- otherwise, set the third motion vector range equal to the first motion vector range,
- during calculation of motion vectors for the affine block perform motion vector clipping
- 5 according to the third range.
- Embodiment 14, wherein the first predefined area is area inside the reference picture.
- Embodiment 15, wherein the first predefined area is area inside the reference picture with a margin, wherein the margin is an extended area surrounding the reference picture.
- Embodiment 16, wherein the margin depends on a maximum CTU size.
- 10 Embodiment 17, wherein the motion vector spread is calculated based on integer numbers predefined for each possible block width and for each possible block height.

- According to another aspect, the disclosure relates to a method for inter-prediction of a current image block in a current picture of a video, the method is used by a
- 15 decoding/encoding device the method comprising:
- calculating a subblock size $M \times N$ based on affine motion model parameters or based on information from which the affine motion model parameters can be derived;
- in the case that either a subblock width M or a subblock height N is smaller than or equal to a predefined value, performing enhanced bi-linear Interpolation Filter (EIF) motion
- 20 compensation process, wherein the performing EIF motion compensation process comprises:
- deriving a motion vector of a respective subblock of an image block (such as an affine image block) based on the affine motion model parameters on a $P \times Q$ (such as 1×1) subblock basis;
- and
- performing clipping on the motion vector of the subblock, so that the clipped motion vector is
- 25 in a motion vector range (such as a second motion vector range). A motion vector range for the coding block is derived based on the center motion vector of the coding block and a size of the coding block, and, if the motion vector range is at least partially located outside a boundary of a reference picture, the method further comprise updating the motion vector range to be located within the picture, such that a starting value or an end value of the
- 30 updated motion vector range is located at the boundary of the reference picture.

FIG. 18 is a block diagram illustrating a decoder 1500 according to the present disclosure. The decoder 1500 comprises one or more processors 1510 and a non-transitory computer-readable storage medium 1520 coupled to the one or more processors 1510 and storing instructions for

execution by the one or more processors 1510, wherein the programming, when executed by the one or more processors 1510, configures the decoder 1500 to carry out the method according to the present disclosure.

FIG. 19 is a block diagram illustrating an encoder 1600 according to the present disclosure.

- 5 The encoder 1600 comprises one or more processors 1610 and a non-transitory computer-readable storage medium 1620 coupled to the one or more processors 1610 and storing instructions for execution by the one or more processors 1610, wherein the programming, when executed by the one or more processors 1610, configures the decoder 1600 to carry out the method according to the present disclosure.

10

Following is an explanation of the applications of the encoding method as well as the decoding method as shown in the above-mentioned embodiments, and a system using them.

- FIG. 15 is a block diagram showing a content supply system 3100 for realizing content
15 distribution service. This content supply system 3100 includes capture device 3102, terminal device 3106, and optionally includes display 3126. The capture device 3102 communicates with the terminal device 3106 over communication link 3104. The communication link may include the communication channel 13 described above. The communication link 3104 includes but not limited to WIFI, Ethernet, Cable, wireless (3G/4G/5G), USB, or any kind of
20 combination thereof, or the like.

- The capture device 3102 generates data, and may encode the data by the encoding method as shown in the above embodiments. Alternatively, the capture device 3102 may distribute the data to a streaming server (not shown in the Figures), and the server encodes the data and transmits the encoded data to the terminal device 3106. The capture device 3102 includes but
25 not limited to camera, smart phone or Pad, computer or laptop, video conference system, PDA, vehicle mounted device, or a combination of any of them, or the like. For example, the capture device 3102 may include the source device 12 as described above. When the data includes video, the video encoder 20 included in the capture device 3102 may actually perform video encoding processing. When the data includes audio (i.e., voice), an audio encoder included in
30 the capture device 3102 may actually perform audio encoding processing. For some practical scenarios, the capture device 3102 distributes the encoded video and audio data by multiplexing them together. For other practical scenarios, for example in the video conference system, the encoded audio data and the encoded video data are not multiplexed. Capture device 3102

distributes the encoded audio data and the encoded video data to the terminal device 3106 separately.

In the content supply system 3100, the terminal device 310 receives and reproduces the encoded data. The terminal device 3106 could be a device with data receiving and recovering capability, such as smart phone or Pad 3108, computer or laptop 3110, network video recorder (NVR)/ digital video recorder (DVR) 3112, TV 3114, set top box (STB) 3116, video conference system 3118, video surveillance system 3120, personal digital assistant (PDA) 3122, vehicle mounted device 3124, or a combination of any of them, or the like capable of decoding the above-mentioned encoded data. For example, the terminal device 3106 may include the destination device 14 as described above. When the encoded data includes video, the video decoder 30 included in the terminal device is prioritized to perform video decoding. When the encoded data includes audio, an audio decoder included in the terminal device is prioritized to perform audio decoding processing.

For a terminal device with its display, for example, smart phone or Pad 3108, computer or laptop 3110, network video recorder (NVR)/ digital video recorder (DVR) 3112, TV 3114, personal digital assistant (PDA) 3122, or vehicle mounted device 3124, the terminal device can feed the decoded data to its display. For a terminal device equipped with no display, for example, STB 3116, video conference system 3118, or video surveillance system 3120, an external display 3126 is contacted therein to receive and show the decoded data.

When each device in this system performs encoding or decoding, the picture encoding device or the picture decoding device, as shown in the above-mentioned embodiments, can be used.

FIG. 16 is a diagram showing a structure of an example of the terminal device 3106. After the terminal device 3106 receives stream from the capture device 3102, the protocol proceeding unit 3202 analyzes the transmission protocol of the stream. The protocol includes but not limited to Real Time Streaming Protocol (RTSP), Hyper Text Transfer Protocol (HTTP), HTTP Live streaming protocol (HLS), MPEG-DASH, Real-time Transport protocol (RTP), Real Time Messaging Protocol (RTMP), or any kind of combination thereof, or the like.

After the protocol proceeding unit 3202 processes the stream, stream file is generated. The file is outputted to a demultiplexing unit 3204. The demultiplexing unit 3204 can separate the multiplexed data into the encoded audio data and the encoded video data. As described above, for some practical scenarios, for example in the video conference system, the encoded audio data and the encoded video data are not multiplexed. In this situation, the encoded data is

transmitted to video decoder 3206 and audio decoder 3208 without through the demultiplexing unit 3204.

Via the demultiplexing processing, video elementary stream (ES), audio ES, and optionally subtitle are generated. The video decoder 3206, which includes the video decoder 30 as explained in the above mentioned embodiments, decodes the video ES by the decoding method as shown in the above-mentioned embodiments to generate video frame, and feeds this data to the synchronous unit 3212. The audio decoder 3208, decodes the audio ES to generate audio frame, and feeds this data to the synchronous unit 3212. Alternatively, the video frame may store in a buffer (not shown in FIG. Y) before feeding it to the synchronous unit 3212. Similarly, the audio frame may store in a buffer before feeding it to the synchronous unit 3212.

The synchronous unit 3212 synchronizes the video frame and the audio frame, and supplies the video/audio to a video/audio display 3214. For example, the synchronous unit 3212 synchronizes the presentation of the video and audio information. Information may code in the syntax using time stamps concerning the presentation of coded audio and visual data and time stamps concerning the delivery of the data stream itself.

If subtitle is included in the stream, the subtitle decoder 3210 decodes the subtitle, and synchronizes it with the video frame and the audio frame, and supplies the video/audio/subtitle to a video/audio/subtitle display 3216.

The present invention is not limited to the above-mentioned system, and either the picture encoding device or the picture decoding device in the above-mentioned embodiments can be incorporated into other system, for example, a car system.

Mathematical Operators

The mathematical operators used in this application are similar to those used in the C programming language. However, the results of integer division and arithmetic shift operations are defined more precisely, and additional operations are defined, such as exponentiation and real-valued division. Numbering and counting conventions generally begin from 0, e.g., "the first" is equivalent to the 0-th, "the second" is equivalent to the 1-th, etc.

Arithmetic operators

The following arithmetic operators are defined as follows:

+ Addition

− Subtraction (as a two-argument operator) or negation (as a unary prefix operator)

*	Multiplication, including matrix multiplication
x^y	Exponentiation. Specifies x to the power of y. In other contexts, such notation is used for superscripting not intended for interpretation as exponentiation.
/	Integer division with truncation of the result toward zero. For example, $7 / 4$ and $-7 / -4$ are truncated to 1 and $-7 / 4$ and $7 / -4$ are truncated to -1.
$\div$	Used to denote division in mathematical equations where no truncation or rounding is intended.
$\frac{x}{y}$	Used to denote division in mathematical equations where no truncation or rounding is intended.
$\sum_{i=x}^y f(i)$	The summation of $f(i)$ with i taking all integer values from x up to and including y.
$x \% y$	Modulus. Remainder of x divided by y, defined only for integers x and y with $x \geq 0$ and $y > 0$.

Logical operators

The following logical operators are defined as follows:

	$x \&\& y$	Boolean logical "and" of x and y
	$x y$	Boolean logical "or" of x and y
5	!	Boolean logical "not"
	$x ? y : z$	If x is TRUE or not equal to 0, evaluates to the value of y; otherwise, evaluates to the value of z.

Relational operators

The following relational operators are defined as follows:

10	>	Greater than
	>=	Greater than or equal to
	<	Less than
	<=	Less than or equal to
	= =	Equal to
15	!=	Not equal to

When a relational operator is applied to a syntax element or variable that has been assigned the value "na" (not applicable), the value "na" is treated as a distinct value for the syntax element or variable. The value "na" is considered not to be equal to any other value.

Bit-wise operators

5 The following bit-wise operators are defined as follows:

- & Bit-wise "and". When operating on integer arguments, operates on a two's complement representation of the integer value. When operating on a binary argument that contains fewer bits than another argument, the shorter argument is extended by adding more significant bits equal to 0.
- 10 | Bit-wise "or". When operating on integer arguments, operates on a two's complement representation of the integer value. When operating on a binary argument that contains fewer bits than another argument, the shorter argument is extended by adding more significant bits equal to 0.
- 15 ^ Bit-wise "exclusive or". When operating on integer arguments, operates on a two's complement representation of the integer value. When operating on a binary argument that contains fewer bits than another argument, the shorter argument is extended by adding more significant bits equal to 0.
- x >> y Arithmetic right shift of a two's complement integer representation of x by y binary digits. This function is defined only for non-negative integer values of y.
- 20 Bits shifted into the most significant bits (MSBs) as a result of the right shift have a value equal to the MSB of x prior to the shift operation.
- x << y Arithmetic left shift of a two's complement integer representation of x by y binary digits. This function is defined only for non-negative integer values of y.
- 25 Bits shifted into the least significant bits (LSBs) as a result of the left shift have a value equal to 0.

Assignment operators

The following arithmetic operators are defined as follows:

- = Assignment operator
- ++ Increment, i.e., $x++$ is equivalent to $x = x + 1$; when used in an array index, evaluates to the value of the variable prior to the increment operation.
- 30 -- Decrement, i.e., $x--$ is equivalent to $x = x - 1$; when used in an array index, evaluates to the value of the variable prior to the decrement operation.

$+=$ Increment by amount specified, i.e., $x += 3$ is equivalent to $x = x + 3$, and $x += (-3)$ is equivalent to $x = x + (-3)$.

$-=$ Decrement by amount specified, i.e., $x -= 3$ is equivalent to $x = x - 3$, and $x -= (-3)$ is equivalent to $x = x - (-3)$.

5 Range notation

The following notation is used to specify a range of values:

$x = y..z$ x takes on integer values starting from y to z , inclusive, with x , y , and z being integer numbers and z being greater than y .

Mathematical functions

10 The following mathematical functions are defined:

$$\text{Abs}(x) = \begin{cases} x & ; \quad x \geq 0 \\ -x & ; \quad x < 0 \end{cases}$$

$\text{Asin}(x)$ the trigonometric inverse sine function, operating on an argument x that is in the range of -1.0 to 1.0 , inclusive, with an output value in the range of $-\pi/2$ to $\pi/2$, inclusive, in units of radians

15 $\text{Atan}(x)$ the trigonometric inverse tangent function, operating on an argument x , with an output value in the range of $-\pi/2$ to $\pi/2$, inclusive, in units of radians

$$\text{Atan2}(y, x) = \begin{cases} \text{Atan}\left(\frac{y}{x}\right) & ; \quad x > 0 \\ \text{Atan}\left(\frac{y}{x}\right) + \pi & ; \quad x < 0 \ \&\& \ y \geq 0 \\ \text{Atan}\left(\frac{y}{x}\right) - \pi & ; \quad x < 0 \ \&\& \ y < 0 \\ +\frac{\pi}{2} & ; \quad x = 0 \ \&\& \ y \geq 0 \\ -\frac{\pi}{2} & ; \quad \text{otherwise} \end{cases}$$

$\text{Ceil}(x)$ the smallest integer greater than or equal to x .

$$\text{Clip1}_Y(x) = \text{Clip3}(0, (1 \ll \text{BitDepth}_Y) - 1, x)$$

20 $\text{Clip1}_C(x) = \text{Clip3}(0, (1 \ll \text{BitDepth}_C) - 1, x)$

$$\text{Clip3}(x, y, z) = \begin{cases} x & ; \quad z < x \\ y & ; \quad z > y \\ z & ; \quad \text{otherwise} \end{cases}$$

$\text{Cos}(x)$ the trigonometric cosine function operating on an argument x in units of radians.

$\text{Floor}(x)$ the largest integer less than or equal to x .

$$\text{GetCurrMsb}(a, b, c, d) = \begin{cases} c + d & ; \quad b - a \geq d / 2 \\ c - d & ; \quad a - b > d / 2 \\ c & ; \quad \text{otherwise} \end{cases}$$

$\text{Ln}(x)$ the natural logarithm of x (the base- e logarithm, where e is the natural logarithm base constant 2.718 281 828...).

$\text{Log2}(x)$ the base-2 logarithm of x .

5 $\text{Log10}(x)$ the base-10 logarithm of x .

$$\text{Min}(x, y) = \begin{cases} x & ; \quad x \leq y \\ y & ; \quad x > y \end{cases}$$

$$\text{Max}(x, y) = \begin{cases} x & ; \quad x \geq y \\ y & ; \quad x < y \end{cases}$$

$$\text{Round}(x) = \text{Sign}(x) * \text{Floor}(\text{Abs}(x) + 0.5)$$

$$\text{Sign}(x) = \begin{cases} 1 & ; \quad x > 0 \\ 0 & ; \quad x == 0 \\ -1 & ; \quad x < 0 \end{cases}$$

10 $\text{Sin}(x)$ the trigonometric sine function operating on an argument x in units of radians

$$\text{Sqrt}(x) = \sqrt{x}$$

$$\text{Swap}(x, y) = (y, x)$$

$\text{Tan}(x)$ the trigonometric tangent function operating on an argument x in units of radians

Order of operation precedence

15 When an order of precedence in an expression is not indicated explicitly by use of parentheses, the following rules apply:

- Operations of a higher precedence are evaluated before any operation of a lower precedence.
- Operations of the same precedence are evaluated sequentially from left to right.

20

The table below specifies the precedence of operations from highest to lowest; a higher position in the table indicates a higher precedence.

25 For those operators that are also used in the C programming language, the order of precedence used in this Specification is the same as used in the C programming language.

Table: Operation precedence from highest (at top of table) to lowest (at bottom of table)

operations (with operands x, y, and z)
"x++", "x--"
"!x", "-x" (as a unary prefix operator)
x^y
"x * y", "x / y", "x ÷ y", " $\frac{x}{y}$ ", "x % y"
"x + y", "x - y" (as a two-argument operator), " $\sum_{i=x}^y f(i)$ "
"x << y", "x >> y"
"x < y", "x <= y", "x > y", "x >= y"
"x == y", "x != y"
"x & y"
"x y"
"x && y"
"x y"
"x ? y : z"
"x.y"
"x = y", "x += y", "x -= y"

Text description of logical operations

In the text, a statement of logical operations as would be described mathematically in the following form:

```

5      if( condition 0 )
          statement 0
      else if( condition 1 )
          statement 1
      ...

```

```

else /* informative remark on remaining condition */
    statement n

```

may be described in the following manner:

... as follows / ... the following applies:

- 5
- If condition 0, statement 0
 - Otherwise, if condition 1, statement 1
 - ...
 - Otherwise (informative remark on remaining condition), statement n

Each "If ... Otherwise, if ... Otherwise, ..." statement in the text is introduced with "... as follows" or "... the following applies" immediately followed by "If ... ". The last condition of the "If ... Otherwise, if ... Otherwise, ..." is always an "Otherwise, ...". Interleaved "If ... Otherwise, if ... Otherwise, ..." statements can be identified by matching "... as follows" or "... the following applies" with the ending "Otherwise, ...".

15 In the text, a statement of logical operations as would be described mathematically in the following form:

```

if( condition 0a  &&  condition 0b )
    statement 0
else if( condition 1a  ||  condition 1b )
    statement 1
20 ...
else
    statement n

```

may be described in the following manner:

... as follows / ... the following applies:

- 25
- If all of the following conditions are true, statement 0:
 - condition 0a
 - condition 0b
 - Otherwise, if one or more of the following conditions are true, statement 1:
 - condition 1a

- condition 1b
- ...
- Otherwise, statement n

In the text, a statement of logical operations as would be described mathematically in the
 5 following form:

```

if( condition 0 )
    statement 0
if( condition 1 )
    statement 1
  
```

10 may be described in the following manner:

When condition 0, statement 0

When condition 1, statement 1.

Although embodiments of the invention have been primarily described based on video coding, it should be noted that embodiments of the coding system 10, encoder 20 and decoder
 15 30 (and correspondingly the system 10) and the other embodiments described herein may also be configured for still picture processing or coding, i.e. the processing or coding of an individual picture independent of any preceding or consecutive picture as in video coding. In general only inter-prediction units 244 (encoder) and 344 (decoder) may not be available in case the picture processing coding is limited to a single picture 17. All other functionalities
 20 (also referred to as tools or technologies) of the video encoder 20 and video decoder 30 may equally be used for still picture processing, e.g. residual calculation 204/304, transform 206, quantization 208, inverse quantization 210/310, (inverse) transform 212/312, partitioning 262/362, intra-prediction 254/354, and/or loop filtering 220, 320, and entropy coding 270 and entropy decoding 304.

25 Embodiments, e.g. of the encoder 20 and the decoder 30, and functions described herein, e.g. with reference to the encoder 20 and the decoder 30, may be implemented in hardware, software, firmware, or any combination thereof. If implemented in software, the functions may be stored on a computer-readable medium or transmitted over communication media as one or more instructions or code and executed by a hardware-based processing unit.

30 Computer-readable media may include computer-readable storage media, which corresponds to a tangible medium such as data storage media, or communication media including any medium that facilitates transfer of a computer program from one place to another, e.g.,

according to a communication protocol. In this manner, computer-readable media generally may correspond to (1) tangible computer-readable storage media which is non-transitory or (2) a communication medium such as a signal or carrier wave. Data storage media may be any available media that can be accessed by one or more computers or one or more
5 processors to retrieve instructions, code and/or data structures for implementation of the techniques described in this disclosure. A computer program product may include a computer-readable medium.

By way of example, and not limiting, such computer-readable storage media can comprise RAM, ROM, EEPROM, CD-ROM or other optical disk storage, magnetic disk storage, or
10 other magnetic storage devices, flash memory, or any other medium that can be used to store desired program code in the form of instructions or data structures and that can be accessed by a computer. Also, any connection is properly termed a computer-readable medium. For example, if instructions are transmitted from a website, server, or other remote source using a coaxial cable, fiber optic cable, twisted pair, digital subscriber line (DSL), or wireless
15 technologies such as infrared, radio, and microwave, then the coaxial cable, fiber optic cable, twisted pair, DSL, or wireless technologies such as infrared, radio, and microwave are included in the definition of medium. It should be understood, however, that computer-readable storage media and data storage media do not include connections, carrier waves, signals, or other transitory media, but are instead directed to non-transitory, tangible storage
20 media. Disk and disc, as used herein, includes compact disc (CD), laser disc, optical disc, digital versatile disc (DVD), floppy disk and Blu-ray disc, where disks usually reproduce data magnetically, while discs reproduce data optically with lasers. Combinations of the above should also be included within the scope of computer-readable media.

Instructions may be executed by one or more processors, such as one or more digital signal
25 processors (DSPs), general purpose microprocessors, application specific integrated circuits (ASICs), field programmable logic arrays (FPGAs), or other equivalent integrated or discrete logic circuitry. Accordingly, the term "processor," as used herein may refer to any of the foregoing structure or any other structure suitable for implementation of the techniques described herein. In addition, in some aspects, the functionality described herein may be
30 provided within dedicated hardware and/or software modules configured for encoding and decoding, or incorporated in a combined codec. Also, the techniques could be fully implemented in one or more circuits or logic elements.

The techniques of this disclosure may be implemented in a wide variety of devices or apparatuses, including a wireless handset, an integrated circuit (IC) or a set of ICs (e.g., a

chip set). Various components, modules, or units are described in this disclosure to emphasize functional aspects of devices configured to perform the disclosed techniques, but do not necessarily require realization by different hardware units. Rather, as described above, various units may be combined in a codec hardware unit or provided by a collection of

5 interoperative hardware units, including one or more processors as described above, in conjunction with suitable software and/or firmware.

CLAIMS1. A method for coding video data, comprising:
obtaining a center motion vector, mv_center of a coding block,
deriving a first motion vector range for the coding block based on the center motion vector
and a motion vector spread, wherein the motion vector spread depends on a size of the coding
5 block;
if the first motion vector range is at least partially pointing outside a first area including a
reference picture, updating the first motion vector range to be pointing within the first area,
such that a minimum value and/or a maximum value of the updated first motion vector range
is pointing at the boundary of the first area, wherein the difference between the maximum
10 value and the minimum value of the updated first motion vector range is equal to the
minimum of a double value of the motion vector spread and the size of the first area; and
performing pixel-based motion compensation based on the updated first motion vector range.

2. The method of claim 1, wherein the motion vector spread is represented by a horizontal
15 motion vector spread and/or a vertical motion vector spread, and the horizontal motion vector
spread is derived based on the width of the coding block, and the vertical motion vector
spread is derived based on the height of the coding block.

3. The method of claim 2, wherein the horizontal motion vector spread is represented by
20 deviationMV[$\log_2\text{CbWidth} - 3$], and the vertical motion vector spread is represented by
deviationMV[$\log_2\text{CbHeight} - 3$], wherein cbWidth and cbHeight represent the width and
the height of the coding block.

4. The method of any one of claims 1 to 3,
25 wherein the first motion vector range is represented by a first MV horizontal component
range and/or a first MV vertical component range,
the first MV horizontal component range comprises a first minimum MV horizontal
component value, hor_min and a first maximum MV horizontal component value, hor_max,
and
30 the first MV vertical component range comprises a first minimum MV vertical component
value, ver_min and a first maximum MV vertical component value, ver_max.

5. The method of claim 4, further comprising:
deriving a second motion vector range based on a size of a picture comprising the coding
block,
wherein the second motion vector range is represented by a second MV horizontal component
5 range and/or a second MV vertical component range,
the second MV horizontal component range comprises a second minimum MV horizontal
component value, hor_min_pic and a second maximum MV horizontal component value,
hor_max_pic and
the second MV vertical component range comprises a second minimum MV vertical
10 component value, ver_min_pic and a second maximum MV vertical component value,
ver_max_pic.

6. The method of claim 5, wherein if the first motion vector range is at least partially pointing
outside a first area including a reference picture, updating the first motion vector range to be
15 pointing within the first area, comprises:
if the first minimum MV horizontal component value, hor_min is less than the second
minimum MV horizontal component value, hor_min_pic,
setting a updated value of the first minimum MV horizontal component value, hor_min equal
to the second minimum MV horizontal component value, hor_min_pic, and
20 deriving a updated value of the first maximum MV horizontal component value, hor_max
based on a sum of the second minimum MV horizontal component value, hor_min_pic and
the double value of the horizontal motion vector spread.

7. The method of claim 5, wherein if the first motion vector range is at least partially pointing
25 outside a first area including a reference picture, updating the first motion vector range to be
pointing within the first area, comprises:
if the first maximum MV horizontal component value, hor_max is greater than the second
maximum MV horizontal component value, hor_max_pic,
setting a updated value of the first maximum MV horizontal component value, hor_max
30 equal to the second maximum MV horizontal component value, hor_max_pic, and
deriving a updated value of the first minimum MV horizontal component value, hor_min
based on a subtraction value of the second maximum MV horizontal component value,
hor_max_pic and the doubled value of the horizontal motion vector spread.

8. The method of any one of claims 1 to 6, wherein
if the first motion vector range is at least partially pointing outside a left boundary of the first
area, updating the variables hor_min and hor_max, representing a first MV horizontal
5 component range of the first motion vector range, as:
hor_min = hor_min_pic,
hor_max = min (hor_max_pic, hor_min_pic + 2* horizontal motion vector spread);
wherein hor_min indicates the updated first minimum MV horizontal component, and
hor_max indicates the updated first maximum MV horizontal component;
10 wherein hor_min_pic and hor_max_pic, representing a second MV horizontal component
range of a second motion vector range;
hor_min_pic indicates a second minimum MV horizontal component of the second MV
horizontal component range, and hor_max_pic indicates a second maximum MV horizontal
component of the second MV horizontal component range, and
15 wherein the second motion vector range depends on a size of a picture comprising the coding
block.

9. The method of any one of claims 1 to 5 and 7, wherein
if the first motion vector range is at least partially pointing outside a right boundary of the
20 first area,
updating the variables hor_min and hor_max, representing a first MV horizontal component
range of the first motion vector range, as:
hor_min = max (hor_min_pic, hor_max_pic – 2* horizontal motion vector spread),
hor_max = hor_max_pic;
25 wherein hor_min indicates the updated first minimum MV horizontal component, and
hor_max indicates the updated first maximum MV horizontal component;
wherein hor_min_pic and hor_max_pic, representing a second MV horizontal component
range of a second motion vector range;
hor_min_pic indicates a second minimum MV horizontal component of the second MV
30 horizontal component range, and hor_max_pic indicates a second maximum MV horizontal
component of the second MV horizontal component range, and
wherein the second motion vector range depends on a size of a picture comprising the coding
block.

10. The method of any one of claims 1 to 5, wherein
if the first motion vector range is at least partially pointing outside an upper boundary of the
first area,
updating the variables ver_min and ver_max, representing a first MV vertical component
5 range of the first motion vector range, as:
ver_min = ver_min_pic,
ver_max = min (ver_max_pic, ver_min_pic + 2* vertical motion vector spread);
wherein ver_min indicates the updated first minimum MV vertical component, and ver_max
indicates the updated first maximum MV vertical component;
10 wherein ver_min_pic and ver_max_pic, representing a second MV vertical component range
of a second motion vector range;
ver_min_pic indicates a second minimum MV vertical component of the second MV vertical
component range, and ver_max_pic indicates a second maximum MV horizontal component
of the second MV horizontal component range, and
15 wherein the second motion vector range depends on a size of a picture comprising the coding
block.

11. The method of any one of claims 1 to 5, wherein
if the first motion vector range is at least partially located outside a lower boundary of the
20 first area,
updating the variables ver_min and ver_max, representing a first MV vertical component
range of the first motion vector range, as:
ver_min = max (ver_min_pic, ver_max_pic – 2* vertical motion vector spread),
ver_max = ver_max_pic;
25 wherein ver_min indicates the updated first minimum MV vertical component of the first MV
vertical component range, and ver_max indicates the updated first maximum MV vertical
component of the first MV vertical component range;
wherein ver_min_pic and ver_max_pic, representing a second MV vertical component range
of a second motion vector range;
30 ver_min_pic indicates a second minimum MV vertical component of the second MV vertical
component range, and ver_max_pic indicates a second maximum MV horizontal component
of the second MV horizontal component range, and
wherein the second motion vector range depends on a size of a picture comprising the coding
block.

12. The method of claim 1, wherein a minimum or maximum value of a horizontal component of the updated first motion vector range is pointing at a left or right boundary of the first area, respectively, and/or wherein a minimum or maximum value of a vertical component of the updated first motion vector range is pointing at an upper or lower boundary of the first area, respectively.

13. The method of any one of claims 1 to 12, wherein the first area comprises the reference picture and an extended area surrounding the reference picture.

10

14. The method of any one of claims 1 to 13, further comprising:
performing a clipping operation on the updated first motion vector range to be within the range $[-2^{17}, 2^{17}-1]$;
wherein the performing pixel-based motion compensation based on the updated first motion vector range, comprising: performing pixel-based motion compensation based on the updated and clipped first motion vector range.

15. The method of any of claims 1 to 14, wherein the performing pixel-based motion compensation based on the updated motion vector range, comprises:
performing clipping operation on a motion vector of a pixel of the coding block to be within a range, to obtain a clipped motion vector, the range depends on the updated first motion vector range; and
performing pixel-based motion compensation based on the clipped motion vector.

16. A coder comprising processing circuitry for carrying out the method according to any one of claims 1 to 15.

25

17. The coder of claim 15, wherein the coder comprises an encoder or a decoder.

18. A computer program product comprising instructions which, when the program is executed by a computer, cause the computer to carry out the method according to any one of the claims 1 to 15.

30

19. A decoder (30, 1800), comprising:
one or more processors (1810); and

a non-transitory computer-readable storage medium (1820) coupled to the one or more processors (1810) and storing instructions for execution by the one or more processors (1810), wherein the programming, when executed by the one or more processors (1810), configures the decoder (30, 1800) to carry out the method according to any one of claims 1 to 15.

5

20. An encoder (20, 1900), comprising:

one or more processors (1910); and

a non-transitory computer-readable storage medium (1920) coupled to the one or more

processors (1910) and storing instructions for execution by the one or more processors

10 (1910), wherein the programming, when executed by the processors (1910), configures the encoder (20, 1900) to carry out the method according to any one of the claims 1 to 15.