(11) Application No. AU 2003204477 B2

(19) AUSTRALIAN PATENT OFFICE

(54) Title

Encoding and decoding video data

(51)[6] International Patent Classification(s)

H04N 7/32 (2006.01) 7/36
20060101ALI2005122
0BMJP **H04N**
7/26
20060101ALI2005100
8BMEP **H04N**
7/46
20060101ALI2005100
8BMEP **H04N**
7/50
20060101ALI2005100
8BMEP

G06T 9/00 (2006.01)

H03M 7/36 (2006.01)

H04N 7/26 (2006.01)

H04N 7/46 (2006.01)

H04N 7/50 (2006.01)

H04N 7/32
20060101AFI2005122
0BMJP **G06T**
9/00
20060101ALI2005100
8BMEP **H03M**

(21) Application No: 2003204477

(22) Application Date: 2003.05.29

(30) Priority Data

(31) Number	(32) Date	(33) Country
60385965	2002.06.03	US

(43) Publication Date : 2004.01.08

(43) Publication Journal Date : 2004.01.08

(71) Applicant(s)

Microsoft Corporation

(72) Inventor(s)

Tourapis, Alexandros; Wu, Feng; Li, Shipeng

(74) Agent/Attorney

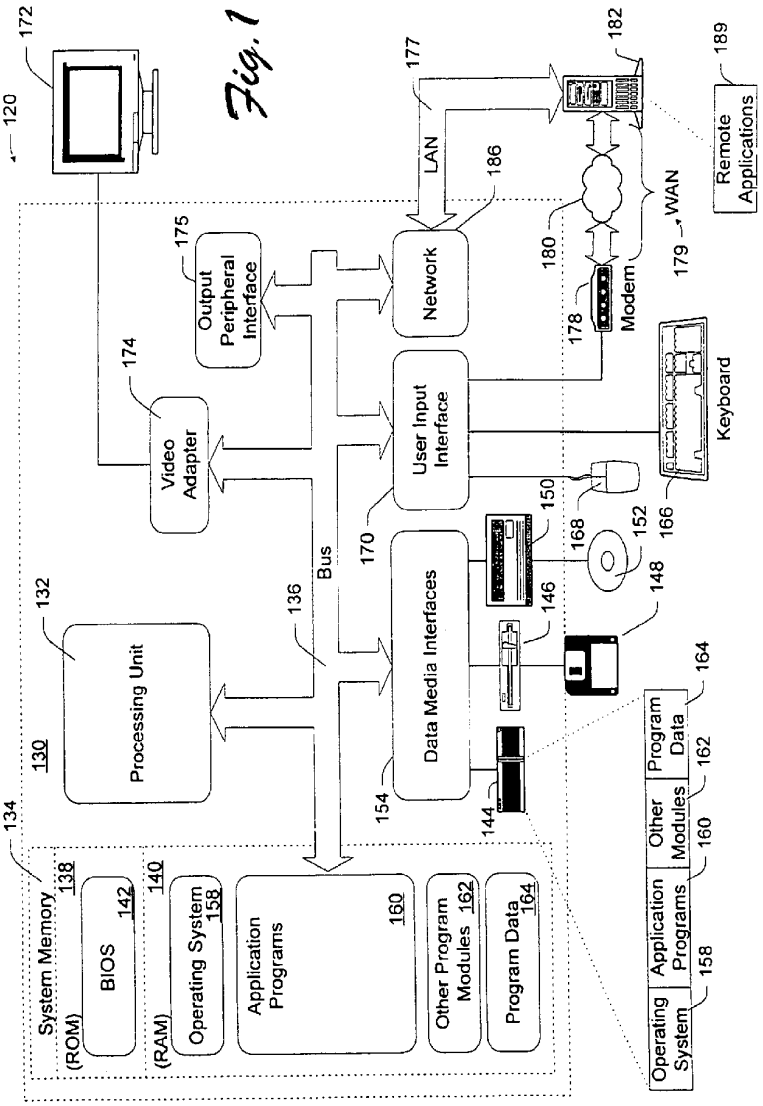
Davies Collison Cave, 1 Nicholson Street, Melbourne, VIC, 3000

(56) Related Art

WO 2002/037859
WO 2001/095633
WO 2002/043399
US 6205177

ABSTRACT

Several improvements for use with Bidirectionally Predictive (B) pictures within a video sequence are provided. In certain improvements Direct Mode encoding and/or Motion Vector Prediction are enhanced using spatial prediction techniques. In other improvements Motion Vector prediction includes temporal distance and subblock information, for example, for more accurate prediction. Such improvements and other presented herein significantly improve the performance of any applicable video coding system/logic.


120
172
Fig. 1
177
182
189
LAN
186
Remote Applications
175
Output Peripheral Interface
Network
180
WAN
179
174
Video Adapter
User Input Interface
178
Modem
Keyboard
132
Processing Unit
136
Bus
170
150
168
152
166
130
Data Media Interfaces
146
148
134
System Memory
138
(ROM)
BIOS
142
140
(RAM)
Operating System
158
Application Programs
160
Other Program Modules
162
Program Data
164
154
144
Operating System
158
Application Programs
160
Other Modules
162
Program Data
164

AUSTRALIA

PATENTS ACT 1990

COMPLETE SPECIFICATION

NAME OF APPLICANT(S)::

Microsoft Corporation

ADDRESS FOR SERVICE:

DAVIES COLLISON CAVE
Patent Attorneys
1 Little Collins Street, Melbourne, 3000.

INVENTION TITLE:

Spatiotemporal prediction for bidirectionally predictive (B) pictures and motion vector prediction for multi-picture reference motion compensation

The following statement is a full description of this invention, including the best method of performing it known to me/us:-

5102

2003204477 29 Jan 2009

This invention relates to encoding and decoding video data.

The motivation for increased coding efficiency in video coding has led to the adoption in the Joint Video Team (JVT) (a standards body) of more refined and complicated models and modes describing motion information for a given macroblock. These models and modes tend to make better advantage of the temporal redundancies that may exist within a video sequence. See, for example, ITU-T, Video Coding Expert Group (VCEG), "JVT Coding – (ITU-T H.26L & ISO/IEC JTC1 Standard) – Working Draft Number 2 (WD-2)", ITU-T JVT-B118, Mar. 2002; and/or Heiko Schwarz and Thomas Wiegand, "Tree-structured macroblock partition", Doc. VCEG-O17, Dec. 2001.

There is continuing need for further improved methods and apparatuses that can support the latest models and modes and also possibly introduce new models and modes to take advantage of improved coding techniques.

It is desired to address one or more of the above limitations, or at least provide a useful alternative.

In accordance with the present invention, there is provided a method for use in encoding video data in a video encoder, the method comprising:

making a spatial/temporal motion vector prediction decision for at least one direct mode macroblock in a B-picture, including deciding between using spatial motion vector prediction for the at least one direct mode macroblock in the B-picture and using temporal motion vector prediction for the at least one direct mode macroblock in the B-picture; and

signalling spatial/temporal motion vector prediction decision information for the at least one direct mode macroblock in a header that includes header information for plural macroblocks in the B-picture, wherein the signalling of the spatial/temporal motion vector prediction decision information in the header communicates to a video decoder the spatial/temporal motion vector prediction decision for the at least one direct macroblock in the B-picture.

The present invention also provides a method for use in decoding video data

2003204477 29 Jan 2009

in a video decoder, the method comprising:

receiving signalled spatial/temporal motion vector prediction decision information for at least one direct mode macroblock in a header that includes header information for plural macroblocks in a B-picture; and

from the signalled spatial/temporal motion vector prediction decision information in the header, determining a spatial/temporal motion vector prediction decision for the at least one direct mode macroblock in the B-picture, including deciding between using spatial motion vector prediction for the at least one direct mode macroblock in the B-picture and using temporal motion vector prediction for the at least one direct mode macroblock in the B-picture.

In certain exemplary implementations the method includes identifying at least a portion of at least one video frame to be a Bidirectionally Predictive (B) picture, and selectively encoding the B picture using at least spatial prediction to encode at least one motion parameter associated with the B picture. The B picture may include a block, a macroblock, a subblock, a slice, or other like portion of the video frame. For example, when a macroblock portion is used, the method produces a Direct Macroblock.

In certain further exemplary implementations, the method further includes employing linear or non-linear motion vector prediction for the B picture based on at least one reference picture that is at least another portion of the video frame. By way of example, in certain implementations, the method employs median motion vector prediction to produce at least one motion vector.

In still other exemplary implementations, in addition to spatial prediction, the method may also process at least one other portion of at least one other video frame to further selectively encode the B picture using temporal prediction to encode at least one temporal-based motion parameter associated with the B picture. In some instances the temporal prediction includes bidirectional temporal prediction, for example based on at least a portion of a Predictive (P) frame.

In certain other implementations, the method also selectively determines

2003204477 29 Jan 2009

applicable scaling for a temporal-based motion parameter based at least in part on a temporal distance between the predictor video frame and the frame that includes the B picture. In certain implementations temporal distance information is encoded, for example, within a header or other like data arrangement associated with the encoded B picture.

Preferred embodiments of the present invention are described hereinafter by

2003204477 23 May 2008

way of example only, with reference to the accompanying drawings. (The same numbers are used throughout the figures to reference like components and/or features.)

Fig. 1 is a block diagram depicting an exemplary computing environment that is suitable for use with certain implementations of the present invention.

Fig. 2 is a block diagram depicting an exemplary representative device that is suitable for use with certain implementations of the present invention.

Fig. 3 is an illustrative diagram depicting spatial predication associated with portions of a picture, in accordance with certain exemplary implementations of the present invention.

Fig. 4 is an illustrative diagram depicting Direct Prediction in B picture coding, in accordance with certain exemplary implementations of the present invention.

Fig. 5 is an illustrative diagram depicting what happens when a scene change happens or even when the collocated block is intra-coded, in accordance with certain exemplary implementations of the present invention.

Fig. 6 is an illustrative diagram depicting handling of collocated intra within existing codecs wherein motion is assumed to be zero, in accordance with certain exemplary implementations of the present invention.

Fig. 7 is an illustrative diagram depicting how Direct Mode is handled when the reference picture of the collocated block in the subsequent P picture is other than zero, in accordance with certain exemplary implementations of the present invention.

Fig. 8 is an illustrative diagram depicting an exemplary scheme wherein MV_{FW} and MV_{BW} are derived from spatial prediction, in accordance with certain exemplary implementations of the present invention.

Fig. 9 is an illustrative diagram depicting how spatial prediction solves the problem of scene changes and the like, in accordance with certain exemplary implementations of the present invention.

Fig. 10 is an illustrative diagram depicting joint spatio-temporal prediction for Direct Mode in B picture coding, in accordance with certain exemplary implementations of the present invention.

Fig. 11 is an illustrative diagram depicting Motion Vector Prediction of a current block considering reference picture information of predictor macroblocks, in accordance with certain exemplary implementations of the present invention.

Fig. 12 is an illustrative diagram depicting how to use more candidates for Direct Mode prediction especially if bidirectional prediction is used within the B picture, in accordance with certain exemplary implementations of the present invention.

Fig. 13 is an illustrative diagram depicting how B pictures may be restricted in using future and past reference pictures, in accordance with certain exemplary implementations of the present invention.

2003204477 23 May 2008

Fig.14 is an illustrative diagram depicting projection of collocated Motion Vectors to a current reference for temporal direct prediction, in accordance with certain exemplary implementations of the present invention.

Figs 15a-c are illustrative diagrams depicting Motion Vector Predictors for one MV in different configurations, in accordance with certain exemplary implementations of the present invention.

Figs 16a-c are illustrative diagrams depicting Motion Vector Predictors for one MV with 8x8 partitions in different configurations, in accordance with certain exemplary implementations of the present invention.

Figs 17a-c are illustrative diagrams depicting Motion Vector Predictors for one MV with additional predictors for 8x8 partitioning, in accordance with certain exemplary implementations of the present invention.

Several improvements for use with Bidirectionally Predictive (B) pictures within a video sequence are described below and illustrated in the accompanying drawings. In certain improvements Direct Mode encoding and/or Motion Vector Prediction are enhanced using spatial prediction techniques. In other improvements Motion Vector prediction includes temporal distance and subblock information, for example, for more accurate prediction. Such improvements and other presented herein significantly improve the performance of any applicable video coding system/logic.

While these and other exemplary methods and apparatuses are described, it should be kept in mind that the techniques of the present invention are not limited to the examples described and shown in the accompanying drawings, but are also clearly adaptable to other similar existing and future video coding schemes, etc.

Before introducing such exemplary methods and apparatuses, an introduction is provided in the following section for suitable exemplary operating environments,

2003204477 23 May 2008

for example, in the form of a computing device and other types of devices/appliances.

Exemplary Operational Environments:

Turning to the drawings, wherein the like reference numerals refer to like elements, preferred embodiments of the invention are illustrated as being implemented in a suitable computing environment, described in the general context of computer-executable instructions, such as program modules, being executed by a personal computer.

Generally, program modules include routines, programs, objects, components, data structures, etc. that perform particular tasks or implement particular abstract data types. Those skilled in the art will appreciate that preferred embodiments of the invention may be practiced with other computer system configurations, including hand-held devices, multi-processor systems, microprocessor based or programmable consumer electronics, network PCs, minicomputers, mainframe computers, portable communication devices, and the like.

Preferred embodiments of the invention may also be practiced in distributed computing environments where tasks are performed by remote processing devices that are linked through a communications network. In a distributed computing environment, program modules may be located in both local and remote memory storage devices.

Fig. 1 illustrates an example of a suitable computing environment 120 on which the subsequently described systems, apparatuses and methods may be implemented. Exemplary computing environment 120 is only one example of a suitable computing environment and is not intended to suggest any limitation as to

the scope of use or functionality of the improved methods and systems described herein. Neither should computing environment 120 be interpreted as having any dependency or requirement relating to any one or combination of components illustrated in computing environment 120.

The improved methods and systems herein are operational with numerous other general purpose or special purpose computing system environments or configurations. Examples of well known computing systems, environments, and/or configurations that may be suitable include, but are not limited to, personal computers, server computers, thin clients, thick clients, hand-held or laptop devices, multiprocessor systems, microprocessor-based systems, set top boxes, programmable consumer electronics, network PCs, minicomputers, mainframe computers, distributed computing environments that include any of the above systems or devices, and the like.

As shown in Fig. 1, computing environment 120 includes a general-purpose computing device in the form of a computer 130. The components of computer 130 may include one or more processors or processing units 132, a system memory 134, and a bus 136 that couples various system components including system memory 134 to processor 132.

Bus 136 represents one or more of any of several types of bus structures, including a memory bus or memory controller, a peripheral bus, an accelerated graphics port, and a processor or local bus using any of a variety of bus architectures. By way of example, and not limitation, such architectures include Industry Standard Architecture (ISA) bus, Micro Channel Architecture (MCA) bus, Enhanced ISA (EISA) bus, Video Electronics Standards Association (VESA) local bus, and Peripheral Component Interconnects (PCI) bus also known as Mezzanine bus.

Computer 130 typically includes a variety of computer readable media. Such media may be any available media that is accessible by computer 130, and it includes both volatile and non-volatile media, removable and non-removable media.

In Fig. 1, system memory 134 includes computer readable media in the form of volatile memory, such as random access memory (RAM) 140, and/or non-volatile memory, such as read only memory (ROM) 138. A basic input/output system (BIOS) 142, containing the basic routines that help to transfer information between elements within computer 130, such as during start-up, is stored in ROM 138. RAM 140 typically contains data and/or program modules that are immediately accessible to and/or presently being operated on by processor 132.

Computer 130 may further include other removable/non-removable, volatile/non-volatile computer storage media. For example, Fig. 1 illustrates a hard disk drive 144 for reading from and writing to a non-removable, non-volatile magnetic media (not shown and typically called a "hard drive"), a magnetic disk drive 146 for reading from and writing to a removable, non-volatile magnetic disk 148 (e.g., a "floppy disk"), and an optical disk drive 150 for reading from or writing to a removable, non-volatile optical disk 152 such as a CD-ROM/R/RW, DVD-ROM/R/RW/+R/RAM or other optical media. Hard disk drive 144, magnetic disk drive 146 and optical disk drive 150 are each connected to bus 136 by one or more interfaces 154.

The drives and associated computer-readable media provide nonvolatile storage of computer readable instructions, data structures, program modules, and other data for computer 130. Although the exemplary environment described herein employs a hard disk, a removable magnetic disk 148 and a removable optical disk 152, it should be appreciated by those skilled in the art that other types of computer readable media which can store data that is accessible by a computer, such as

magnetic cassettes, flash memory cards, digital video disks, random access memories (RAMs), read only memories (ROM), and the like, may also be used in the exemplary operating environment.

A number of program modules may be stored on the hard disk, magnetic disk 148, optical disk 152, ROM 138, or RAM 140, including, e.g., an operating system 158, one or more application programs 160, other program modules 162, and program data 164.

The improved methods and systems described herein may be implemented within operating system 158, one or more application programs 160, other program modules 162, and/or program data 164.

A user may provide commands and information into computer 130 through input devices such as keyboard 166 and pointing device 168 (such as a "mouse"). Other input devices (not shown) may include a microphone, joystick, game pad, satellite dish, serial port, scanner, camera, etc. These and other input devices are connected to the processing unit 132 through a user input interface 170 that is coupled to bus 136, but may be connected by other interface and bus structures, such as a parallel port, game port, or a universal serial bus (USB).

A monitor 172 or other type of display device is also connected to bus 136 via an interface, such as a video adapter 174. In addition to monitor 172, personal computers typically include other peripheral output devices (not shown), such as speakers and printers, which may be connected through output peripheral interface 175.

Computer 130 may operate in a networked environment using logical connections to one or more remote computers, such as a remote computer 182. Remote computer 182 may include many or all of the elements and features described herein relative to computer 130.

Logical connections shown in Fig. 1 are a local area network (LAN) 177 and a general wide area network (WAN) 179. Such networking environments are commonplace in offices, enterprise-wide computer networks, intranets, and the Internet.

When used in a LAN networking environment, computer 130 is connected to LAN 177 via network interface or adapter 186. When used in a WAN networking environment, the computer typically includes a modem 178 or other means for establishing communications over WAN 179. Modem 178, which may be internal or external, may be connected to system bus 136 via the user input interface 170 or other appropriate mechanism.

Depicted in Fig. 1, is a specific implementation of a WAN via the Internet. Here, computer 130 employs modem 178 to establish communications with at least one remote computer 182 via the Internet 180.

In a networked environment, program modules depicted relative to computer 130, or portions thereof, may be stored in a remote memory storage device. Thus, e.g., as depicted in Fig. 1, remote application programs 189 may reside on a memory device of remote computer 182. It will be appreciated that the network connections shown and described are exemplary and other means of establishing a communications link between the computers may be used.

Attention is now drawn to Fig. 2, which is a block diagram depicting another exemplary device 200 that is also capable of benefiting from the methods and apparatuses disclosed herein. Device 200 is representative of any one or more devices or appliances that are operatively configured to process video and/or any related types of data in accordance with all or part of the methods and apparatuses described herein and their equivalents. Thus, device 200 may take the form of a computing device as in Fig.1, or some other form, such as, for example, a wireless

device, a portable communication device, a personal digital assistant, a video player, a television, a DVD player, a CD player, a karaoke machine, a kiosk, a digital video projector, a flat panel video display mechanism, a set-top box, a video game machine, etc. In this example, device 200 includes logic 202 configured to process video data, a video data source 204 configured to provide vide data to logic 202, and at least one display module 206 capable of displaying at least a portion of the video data for a user to view. Logic 202 is representative of hardware, firmware, software and/or any combination thereof. In certain implementations, for example, logic 202 includes a compressor/decompressor (codec), or the like. Video data source 204 is representative of any mechanism that can provide, communicate, output, and/or at least momentarily store video data suitable for processing by logic 202. Video reproduction source is illustratively shown as being within and/or without device 200. Display module 206 is representative of any mechanism that a user might view directly or indirectly and see the visual results of video data presented thereon. Additionally, in certain implementations, device 200 may also include some form or capability for reproducing or otherwise handling audio data associated with the video data. Thus, an audio reproduction module 208 is shown.

With the examples of Figs 1 and 2 in mind, and others like them, the next sections focus on certain exemplary methods and apparatuses that may be at least partially practiced using with such environments and with such devices.

Encoding Bidirectionally Predictive (B) Pictures And Motion Vector Prediction

This section describes several exemplary improvements that can be implemented to encode Bidirectionally Predictive (B) pictures and Motion Vector prediction within a video coding system or the like. The exemplary methods and apparatuses can be applied to predict motion vectors and enhancements in the

design of a B picture Direct Mode. Such methods and apparatuses are particularly suitable for multiple picture reference codecs, such as, for example, JVT, and can achieve considerable coding gains especially for panning sequences or scene changes.

Bidirectionally Predictive (B) pictures are an important part of most video coding standards and systems since they tend to increase the coding efficiency of such systems, for example, when compared to only using Predictive (P) pictures. This improvement in coding efficiency is mainly achieved by the consideration of bidirectional motion compensation, which can effectively improve motion compensated prediction and thus allow the encoding of significantly reduced residue information. Furthermore, the introduction of the Direct Prediction mode for a Macroblock/block within such pictures can further increase efficiency considerably (e.g., more than 10-20%) since no motion information is encoded. Such may be accomplished, for example, by allowing the prediction of both forward and backward motion information to be derived directly from the motion vectors used in the corresponding macroblock of a subsequent reference picture.

By way of example, Fig. 4 illustrates Direct Prediction in B picture at time *t+1* coding based on P frames at times *t* and *t+2*, and the applicable motion vectors (MVs). Here, an assumption is made that an object in the picture is moving with constant speed. This makes it possible to predict a current position inside a B picture without having to transmit any motion vectors. The motion vectors $(\overrightarrow{MV}_{fw}, \overrightarrow{MV}_{bw})$ of the Direct Mode versus the motion vector $\overrightarrow{MV}$ of the collocated MB in the first subsequent P reference picture are basically calculated by:

$$\overrightarrow{MV}_{fw} = \frac{TR_B \cdot \overrightarrow{MV}}{TR_D} \text{ and } \overrightarrow{MV}_{bw} = \frac{(TR_B - TR_D) \cdot \overrightarrow{MV}}{TR_D},$$

where TR_B is the temporal distance between the current B picture and the reference picture pointed by the forward MV of the collocated MB, and TR_D is the temporal distance between the future reference picture and the reference picture pointed by the forward MV of the collocated MB.

Unfortunately there are several cases where the existing Direct Mode does not provide an adequate solution, thus not efficiently exploiting the properties of this mode. In particular, existing designs of this mode usually force the motion parameters of the Direct Macroblock, in the case of the collocated Macroblock in the subsequent P picture being Intra coded, to be zero. For example, see Fig. 6, which illustrates handling of collocated intra within existing codecs wherein motion is assumed to be zero. This essentially means that, for this case, the B picture Macroblock will be coded as the average of the two collocated Macroblocks in the first subsequent and past P references. This immediately raises the following concern; if a Macroblock is Intra-coded, then how does one know how much relationship it has with the collocated Macroblock of its reference picture. In some situations, there may be little if any actual relationship. Hence, it is possible that the coding efficiency of the Direct Mode may be reduced. An extreme case can be seen in the case of a scene change as illustrated in Fig. 5. Fig. 5 illustrates what happens when a scene change occurs in the video sequence and/or what happens when the collocated block is intra. Here, in this example, obviously no relationship exists between the two reference pictures given the scene change. In such a case bidirectional prediction would provide little if any benefit. As such, the Direct Mode could be completely wasted. Unfortunately, conventional implementations of the Direct Mode restrict it to always perform a bidirectional prediction of a Macroblock.

Fig. 7 is an illustrative diagram depicting how Direct Mode is handled when the reference picture of the collocated block in the subsequent P picture is other than zero, in accordance with certain implementations of the present invention.

An additional issue with the Direct Mode Macroblocks exists when multi-picture reference motion compensation is used. Until recently, for example, the JVT standard provided the timing distance information (TR_B and TR_D), thus allowing for the proper scaling of the parameters. Recently, this was changed in the new revision of the codec (see, e.g., Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, "Joint Committee Draft (CD) of Joint Video Specification (ITU-T Rec. H.264 | ISO/IEC 14496-10 AVC)", ITU-T JVT-C167, May. 2002, which is incorporated herein by reference). In the new revision, the motion vector parameters of the subsequent P picture are to be scaled equally for the Direct Mode prediction, without taking in account the reference picture information. This could lead to significant performance degradation of the Direct Mode, since the constant motion assumption is no longer followed.

Nevertheless, even if the temporal distance parameters were available, it is not always certain that the usage of the Direct Mode as defined previously is the most appropriate solution. In particular for the B pictures which are closer to a first forward reference picture, the correlation might be much stronger with that picture, than the subsequent reference picture. An extreme example which could contain such cases could be a sequence where *scene A* changes to *scene B*, and then moves back to *scene A* (e.g., as may happen in a news bulletin, etc.). All the above could deter the performance of B picture encoding considerably since Direct Mode will not be effectively exploited within the encoding process.

With these and other concerns in mind, unlike the previous definitions of the Direct Mode where only temporal prediction was used, in accordance with certain

aspects of the present invention, a new Direct Macroblock type is introduced wherein both temporal prediction and/or spatial prediction is considered. The type(s) of prediction used can depend on the type of reference picture information of the first subsequent P reference picture, for example.

In accordance with certain other aspects of the present invention, one may also further considerably improve motion vector prediction for both P and B pictures when multiple picture references are used, by taking in consideration temporal distances, if such are available.

These enhancements are implemented in certain exemplary methods and apparatuses as described below. The methods and apparatuses can achieve significant bitrate reductions while achieving similar or better quality.

Direct Mode Enhancements:

In most conventional video coding systems, Direct Mode is designed as a bidirectional prediction scheme where motion parameters are always predicted in a temporal way from the motion parameters in the subsequent P images. In this section, an enhanced Direct Mode technique is provided in which spatial information may also/alternatively be considered for such predictions.

One or more of the following exemplary techniques may be implemented as needed, for example, depending on the complexity and/or specifications of the system.

One technique is to implement spatial prediction of the motion vector parameters of the Direct Mode without considering temporal prediction. Spatial prediction can be accomplished, for example, using existing Motion Vector prediction techniques used for motion vector encoding (such as, e.g., median prediction). If multiple picture references are used, then the reference picture of the

adjacent blocks may also be considered (even though there is no such restriction and the same reference, e.g. 0, could always be used).

Motion parameters and reference pictures could be predicted as follows and with reference to Fig. 3, which illustrates spatial predication associated with portions A-E (e.g., macroblocks, slices, etc.) assumed to be available and part of a picture. Here, E is predicted in general from A, B, C as Median (A,B,C). If C is actually outside of the picture then D is used instead. If B,C, and D are outside of picture, then only A is used, where as if A does not exist, such is replaced with (0,0). Those skilled in the art will recognize that spatial prediction may be done at a subblock level as well.

In general spatial prediction can be seen as a linear or nonlinear function of all available motion information calculated within a picture or a group of macroblocks/blocks within the same picture.

There are various methods available that may be arranged to predict the reference picture for Direct Mode. For example, one method may be to select a minimum reference picture among the predictions. In another method, a median reference picture may be selected. In certain methods, a selection may be made between a minimum reference picture and median reference picture, e.g., if the minimum is zero. In still other implementations, a higher priority could also be given to either vertical or horizontal predictors (A and B) due to their possibly stronger correlation with E.

If one of the predictions does not exist (e.g., all surrounding macroblocks are predicted with the same direction FW or BW only or are intra), then the existing one is only used (single direction prediction) or such could be predicted from the one available. For example if forward prediction is available then:

$$\overrightarrow{MV}_{bw} = \frac{(TR_B - TR_D) \cdot \overrightarrow{MV}_{fw}}{TR_B}$$

Temporal prediction is used for Macroblocks if the subsequent P reference is non intra as in existing codecs. Attention is now drawn to Fig. 8, in which MV_{FW} and MV_{BW} are derived from spatial prediction (Median MV of surrounding Macroblocks). If either one is not available (i.e., no predictors) then one-direction is used. If a subsequent P reference is intra, then spatial prediction can be used instead as described above. Assuming that no restrictions exist, if one of the predictions is not available then Direct Mode becomes a single direction prediction mode.

This could considerably benefit video coding when the scene changes, for example, as illustrated in Fig. 9, and/or even when fading exists within a video sequence. As illustrated in Fig. 9, spatial prediction may be used to solve the problem of a scene change.

If temporal distance information is not available within a codec, temporal prediction will not be as efficient in the direct mode for blocks when the collocated P reference block has a non-zero reference picture. In such a case, spatial prediction may also be used as above. As an alternative, one may estimate scaling parameters if one of the surrounding macroblocks also uses the same reference picture as the collocated P reference block. Furthermore, special handling may be provided for the case of zero motion (or close to zero motion) with a non-zero reference. Here, regardless of temporal distance forward and backward motion vectors could always be taken as zero. The best solution, however, may be to always examine the reference picture information of surrounding macroblocks and based thereon decide on how the direct mode should be handled in such a case.

More particularly, for example, given a non-zero reference, the following sub cases may be considered:

Case A: Temporal prediction is used if the motion vectors of the collocated P block are zero.

Case B: If all surrounding macroblocks use different reference pictures than the collocated P reference, then spatial prediction appears to be a better choice and temporal prediction is not used.

Case C: If motion flow inside the B picture appears to be quite different than the one in the P reference picture, then spatial prediction is used instead.

Case D: Spatial or temporal prediction of Direct Mode macroblocks could be signaled inside the image header. A pre-analysis of the image could be performed to decide which should be used.

Case E: Correction of the temporally predicted parameters based on spatial information (or vice versa). Thus, for example, if both appear to have the same or approximately the same phase information then the spatial information could be a very good candidate for the direct mode prediction. A correction could also be done on the phase, thus correcting the sub pixel accuracy of the prediction.

Fig. 10 illustrates a joint spatio-temporal prediction for Direct Mode in B picture coding. Here, in this example, Direct Mode can be a 1- to 4-direction mode depending on information available. Instead of using Bi-directional prediction for Direct Mode macroblocks, a multi-hypothesis extension of such mode can be done and multiple predictions used instead.

Combined with the discussion above, Direct Mode macroblocks can be predicted using from one up to four possible motion vectors depending on the

information available. Such can be decided, for example, based on the mode of the collocated P reference image macroblock and on the surrounding macroblocks in the current B picture. In such a case, if the spatial prediction is too different than the temporal one, one of them could be selected as the only prediction in favor of the other. Since spatial prediction as described previously, might favor a different reference picture than the temporal one, the same macroblock might be predicted from more than 2 reference pictures.

The JVT standard does not restrict the first future reference to be a P picture. Hence, in such a standard, a picture can be a B as illustrated in Fig. 12, or even a Multi-Hypothesis (MH) picture. This implies that more motion vectors are assigned per macroblock. This means that one may also use this property to increase the efficiency of the Direct Mode by more effectively exploiting the additional motion information.

In Fig. 12, the first subsequent reference picture is a B picture (pictures B_8 and B_9). This enables one to use more candidates for Direct Mode prediction especially if bidirectional prediction is used within the B picture.

In particular one may perform the following:

a.) If the collocated reference block in the first future reference is using bidirectional prediction, the corresponding motion vectors (forward or backward) are used for calculating the motion vectors of the current block. Since the backward motion vector of the reference corresponds to a future reference picture, special care should be taken in the estimate of the current motion parameters. Attention is drawn, for example to Fig. 12 in which the first subsequent reference picture is a B picture (pictures B_8 and B_9). This enables one to use more candidates for Direct Mode prediction especially if bidirectional prediction is used within the B picture. Thus, as illustrated, the

backward motion vector of B_8 $\overrightarrow{MV}_{B8bw}$ can be calculated as $2\times\overrightarrow{MV}_{B7bw}$ due to the temporal distance between B_8, B_7 and P_6. Similarly for B_9 the backward motion vector can be taken as $\overrightarrow{MV}_{B7bw}$, if though these refer to the B_7. One may also restrict these to refer to the first subsequent P picture, in which case these motion vectors can be scaled accordingly. A similar conclusion can be deduced about the forward motion vectors. Multiple picture reference or intra macroblocks can be handled similar to the previous discussion.

b.) If bidirectional prediction for the collocated block is used, then, in this example, one may estimate four possible predictions for one macroblock for the direct mode case by projecting and inverting the backward and forward motion vectors of the reference.

c.) Selective projection and inversion may be used depending on temporal distance. According to this solution, one selects the motion vectors from the reference picture which are more reliable for the prediction. For example, considering the illustration in Fig. 12, one will note that B_8 is much closer to P_2 than P_6. This implies that the backward motion vector of B_7 may not be a very reliable prediction. In this case, direct mode motion vectors can therefore be calculated only from the forward prediction of B_7. For B_9, however, both motion vectors seem to be adequate enough for the prediction and therefore may be used. Such decisions/information may also be decided/supported within the header of the image. Other conditions and rules may also be implemented. For example, additional spatial confidence of a prediction and/or a motion vector phase may be considered. Note, in particular, that if the forward and backward motion vectors have no relationship, then the backward motion vector might be too unreliable to use.

Single Picture Reference for B Pictures:

A special case exists with the usage of only one picture reference for B pictures (although, typically a forward and a backward reference are necessary) regardless of how many reference pictures are used in P pictures. From observations of encoding sequences in the current JVT codec, for example, it was noted that, if one compares the single-picture reference versus the multi-picture reference case using B pictures, even though encoding performance of P pictures for the multi-picture case is almost always superior to that of the single-picture, the some is not always true for B pictures.

One reason for this observation is the overhead of the reference picture used for each macroblock. Considering that B pictures rely more on motion information than P pictures, the reference picture information overhead reduces the number of bits that are transmitted for the residue information at a given bitrate, which thereby reduces efficiency. A rather easy and efficient solution could be the selection of only one picture reference for either backward or forward motion compensation, thus not needing to transmit any reference picture information.

This is considered with reference to Figs 13 and 14. As illustrated in Fig. 13, B pictures can be restricted in using only one future and past reference pictures. Thus, for direct mode motion vector calculation, projection of the motion vectors is necessary. A projection of the collocated MVs to the current reference for temporal direct prediction is illustrated in Fig. 14 (note that it is possible that $TD_{D,0} > TD_{D,1}$). Thus, in this example, Direct Mode motion parameters are calculated by projecting motion vectors that refer to other reference pictures to the two reference pictures, or by using spatial prediction as in Fig. 13. Note that such options not only allow for possible reduced encoding complexity of B pictures, but also tend to reduce

memory requirements since fewer B pictures (e.g., maximum two) are needed to be stored if B pictures are allowed to reference B pictures.

In certain cases a reference picture of the first future reference picture may no longer be available in the reference buffer. This could immediately generate a problem for the estimate of Direct Mode macroblocks and special handling of such cases is required. Obviously there is no such problem if a single picture reference is used. However, if multiple picture references are desired, then possible solutions include projecting the motion vector(s) to either the first forward reference picture, and/or to the reference picture that was closest to the non available picture. Either solution could be viable, whereas again spatial prediction could be an alternative solution.

Refinements of the motion vector prediction for single- and multi-picture reference motion compensation

Motion vector prediction for multi-picture reference motion compensation can significantly affect the performance of both B and P picture coding. Existing standards, such as, for example, JVT, do not always consider the reference pictures of the macroblocks used in the prediction. The only consideration such standards do make is when only one of the prediction macroblocks uses the same reference. In such a case, only that predictor is used for the motion prediction. There is no consideration of the reference picture if only one or all predictors are using a different reference.

In such a case, for example, and in accordance with certain further aspects of the present invention, one can scale the predictors according to their temporal distance versus the current reference. Attention is drawn to Fig. 11, which illustrates Motion Vector prediction of a current block (C) considering the reference

picture information of predictor macroblocks (Pr) and performance of proper adjustments (e.g., scaling of the predictors).

If predictors A, B, and C use reference pictures with temporal distance TR_A, TR_B, and TR_C respectively, and the current reference picture has a temporal distance equal to TR, then the median predictor is calculated as follows:

$$\overrightarrow{MV}_{pred} = TR \times Median\left(\frac{\overrightarrow{MV}_A}{TR_A}, \frac{\overrightarrow{MV}_B}{TR_B}, \frac{\overrightarrow{MV}_C}{TR_C}\right)$$

If integer computation is to be used, it may be easier to place the multiplication inside the median, thus increasing accuracy. The division could also be replaced with shifting, but that reduces the performance, whereas it might be necessary to handle signed shifting as well (-1>>N = -1). It is thus very important in such cases to have the temporal distance information available for performing the appropriate scaling. Such could also be available within the header, if not predictable otherwise.

Motion Vector prediction as discussed previously is basically median biased, meaning that the median value among a set of predictors is selected for the prediction. If one only uses one type of macroblock (e.g., 16×16) with one Motion Vector (MV), then these predictors can be defined, for example, as illustrated in Fig. 15. Here, MV predictors are shown for one MV. In Fig. 15a, the MB is not in the first row or the last column. In Fig. 15b, the MB is in the last column. In Fig. 15c, the MB is in the first row.

The JVT standard improves on this further by also considering the case that only one of the three predictors exists (i.e. Macroblocks are intra or are using a different reference picture in the case of multi-picture prediction). In such a case, only the existing or same reference predictor is used for the prediction and all others are not examined.

Intra coding does not always imply that a new object has appeared or that scene changes. It might instead, for example, be the case that motion estimation and compensation is inadequate to represent the current object (e.g., search range, motion estimation algorithm used, quantization of residue, etc) and that better results could be achieved through Intra Coding instead. The available motion predictors could still be adequate enough to provide a good motion vector predictor solution.

What is intriguing is the consideration of subblocks within a Macroblock, with each one being assigned different motion information. MPEG-4 and H.263 standards, for example, can have up to four such subblocks (e.g., with size 8x8), where as the JVT standard allows up to sixteen subblocks while also being able to handle variable block sizes (e.g., 4×4, 4×8, 8×4,8×8, 8×16, 16×8, and 16×16). In addition JVT also allows for 8x8 Intra subblocks, thus complicating things even further.

Considering the common cases of JVT and MPEG-4/H.263 (8×8 and 16×16), the predictor set for a 16×16 macroblock is illustrated in Figs 16a-c having a similar arrangement to Figs 15a-c, respectively. Here, Motion Vector predictors are shown for one MV with 8×8 partitions. Even though the described predictors could give reasonable results in some cases, it appears that they may not adequately cover all possible predictions.

Attention is drawn next to Figs 17a-c, which are also in a similar arrangement to Figs 15a-c, respectively. Here, in Figs 17a-c there are two additional predictors that could also be considered in the prediction phase (C_1 and A_2). If 4×4 blocks are also considered, this increases the possible predictors by four.

Instead of employing a median of the three predictors A, B, and C (or A_1, B, and C_2) one may now have some additional, and apparently more reliable, options. Thus, for example, one can observe that predictors A_1, and C_2 are essentially too close with one another and it may be the case that they may not be too representative in the prediction phase. Instead, selecting predictors A_1, C_1, and B seems to be a more reliable solution due to their separation. An alternative could also be the selection of A_2 instead of A_1 but that may again be too close to predictor B. Simulations suggest that the first case is usually a better choice. For the last column A_2 could be used instead of A_1. For the first row either one of A_1 and A_2 or even their average value could be used. Gain up to 1% was noted within JVT with this implementation.

The previous case adds some tests for the last column. By examining Fig17b, for example, it is obvious that such tends to provide the best partitioning available. Thus, an optional solution could be the selection of A_2, C_1, and B (from the upper-left position). This may not always be recommended however, since such an implementation may adversely affect the performance of right predictors.

An alternative solution would be the usage of averages of predictors within a Macroblock. The median may then be performed as follows:

$$\overrightarrow{MV}_{pred} = Median\left(Ave(\overrightarrow{MV}_{C_1}, \overrightarrow{MV}_{C_2}), Ave(\overrightarrow{MV}_{A_1}, \overrightarrow{MV}_{A_2}), \overrightarrow{MV}_B\right).$$

For median row/column calculation, the median can be calculated as:

$$\overrightarrow{MV}_{pred} = Median(Median(\overrightarrow{MV}_{C_1}, \overrightarrow{MV}_{C_2}, \overrightarrow{MV}_D), \cdots$$
$$Median(\overrightarrow{MV}_D, \overrightarrow{MV}_{A_1}, \overrightarrow{MV}_{C_2}), Median(\overrightarrow{MV}_B, \overrightarrow{MV}_{A_1}, \overrightarrow{MV}_{A_2}))$$

Another possible solution is a Median5 solution. This is probably the most complicated solution due to computation (quick-sort or bubble-sort could for

example be used), but could potentially yield the best results. If 4×4 blocks are considered, for example, then Median9 could also be used:

$$\overrightarrow{MV}_{pred} = Median\left(\overrightarrow{MV}_{C_1}, \overrightarrow{MV}_{C_2}, \overrightarrow{MV}_{D}, \overrightarrow{MV}_{B}, \overrightarrow{MV}_{A_1}, \overrightarrow{MV}_{A_2}\right)$$

Considering that JVT allows the existence of Intra subblocks within an Inter Macroblock (e.g., tree macroblock structure), such could also be taken in consideration within the Motion Prediction. If a subblock (e.g., from Macroblocks above or left only) to be used for the MV prediction is Intra, then the adjacent subblock may be used instead. Thus, if A_1 is intra but A_2 is not, then A_1 can be replaced by A_2 in the prediction. A further possibility is to replace one missing Intra Macroblock with the MV predictor from the upper-left position. In Fig. 17a, for example, if C_1 is missing then D may be used instead.

In the above sections, several improvements on B picture Direct Mode and on Motion Vector Prediction were presented. It was illustrated that spatial prediction can also be used for Direct Mode macroblocks; where as Motion Vector prediction should consider temporal distance and subblock information for more accurate prediction. Such considerations should significantly improve the performance of any applicable video coding system.

Conclusion

Although the description above uses language that is specific to structural features and/or methodological acts, it is to be understood that the invention defined in the appended claims is not limited to the specific features or acts described. Rather, the specific features and acts are disclosed as exemplary forms of implementing the invention.

Throughout this specification and the claims which follow, unless the context requires otherwise, the word "comprise", and variations such as "comprises" and "comprising", will be understood to imply the inclusion of a stated integer or step or group of integers or steps but not the exclusion of any other integer or step or group of integers or steps.

The reference to any prior art in this specification is not, and should not be taken as, an acknowledgement or any form of suggestion that that prior art forms part of the common general knowledge in Australia.

2003204477 29 Jan 2009

THE CLAIMS DEFINING THE INVENTION ARE AS FOLLOWS:

1. A method for use in encoding video data in a video encoder, the method comprising:

making a spatial/temporal motion vector prediction decision for at least one direct mode macroblock in a B-picture, including deciding between using spatial motion vector prediction for the at least one direct mode macroblock in the B-picture and using temporal motion vector prediction for the at least one direct mode macroblock in the B-picture; and

signalling spatial/temporal motion vector prediction decision information for the at least one direct mode macroblock in a header that includes header information for plural macroblocks in the B-picture, wherein the signalling of the spatial/temporal motion vector prediction decision information in the header communicates to a video decoder the spatial/temporal motion vector prediction decision for the at least one direct macroblock in the B-picture.

2. The method of claim 1 wherein the plural macroblocks in the B-picture are in a slice of the B-picture.

3. The method of claim 1 wherein the at least one direct mode macroblock comprises plural direct mode macroblocks.

4. The method of claim 3 wherein the plural direct mode macroblocks are 16x16 macroblocks.

5. The method of claim 4 wherein each of the 16x16 macroblocks includes four 8x8 sub-blocks.

6. The method of claim 1 wherein the spatial/temporal motion vector prediction

2003204477 29 Jan 2009

decision for the at least one direct mode macroblock indicates spatial motion vector prediction for the at least one direct mode macroblock, the method further comprising, for a given direct mode macroblock of the at least one direct mode macroblock, selecting a reference picture for the given direct mode macroblock from among reference pictures used for plural surrounding portions.

7. The method of claim 6 wherein the act of selecting the reference picture for the given direct mode macroblock comprises selecting a minimum reference picture for the given direct mode macroblock from among at least two reference pictures used for the surrounding portions.

8. The method of claim 6 wherein the plural surrounding portions are plural surrounding macroblocks.

9. The method of claim 1 wherein the spatial/temporal motion vector prediction decision for the at least one direct mode macroblock indicates using the spatial motion vector prediction for the at least one direct mode macroblock, the spatial motion vector prediction comprises median motion vector prediction.

10. A method for use in decoding video data in a video decoder, the method comprising:

receiving signalled spatial/temporal motion vector prediction decision information for at least one direct mode macroblock in a header that includes header information for plural macroblocks in a B-picture; and

from the signalled spatial/temporal motion vector prediction decision information in the header, determining a spatial/temporal motion vector prediction decision for the at least one direct mode macroblock in the B-picture, including deciding between using spatial motion vector predicton for the at least one direct mode macroblock in the B-picture and using temporal motion vector prediction for

2003204477 29 Jan 2009

the at least one direct mode macroblock in the B-picture.

11. The method of claim 10 wherein the plural macroblocks in the B-picture are in a slice of the B-picture.

12. The method of claim 10 wherein the at least one direct mode macroblock comprises plural direct mode macroblocks.

13. The method of claim 12 wherein the plural direct mode macroblocks are 16x16 macroblocks.

14. The method of claim 13 wherein each of the 16x16 macroblocks includes four 8x8 sub-blocks.

15. The method of claim 10 wherein the spatial/temporal motion vector prediction decision for the at least one direct mode macroblock indicates using the spatial motion vector prediction for the at least one direct mode macroblock, the method further comprising, for a given direct mode macroblock of the at least one direct mode macroblock, selecting a reference picture for the given direct mode macroblock from among reference pictures used for plural surrounding portions.

16. The method of claim 15 wherein the act of selecting the reference picture for the given direct mode macroblock comprises selecting a minimum reference picture for the given direct mode macroblock from among at least two reference pictures used for the surrounding portions.

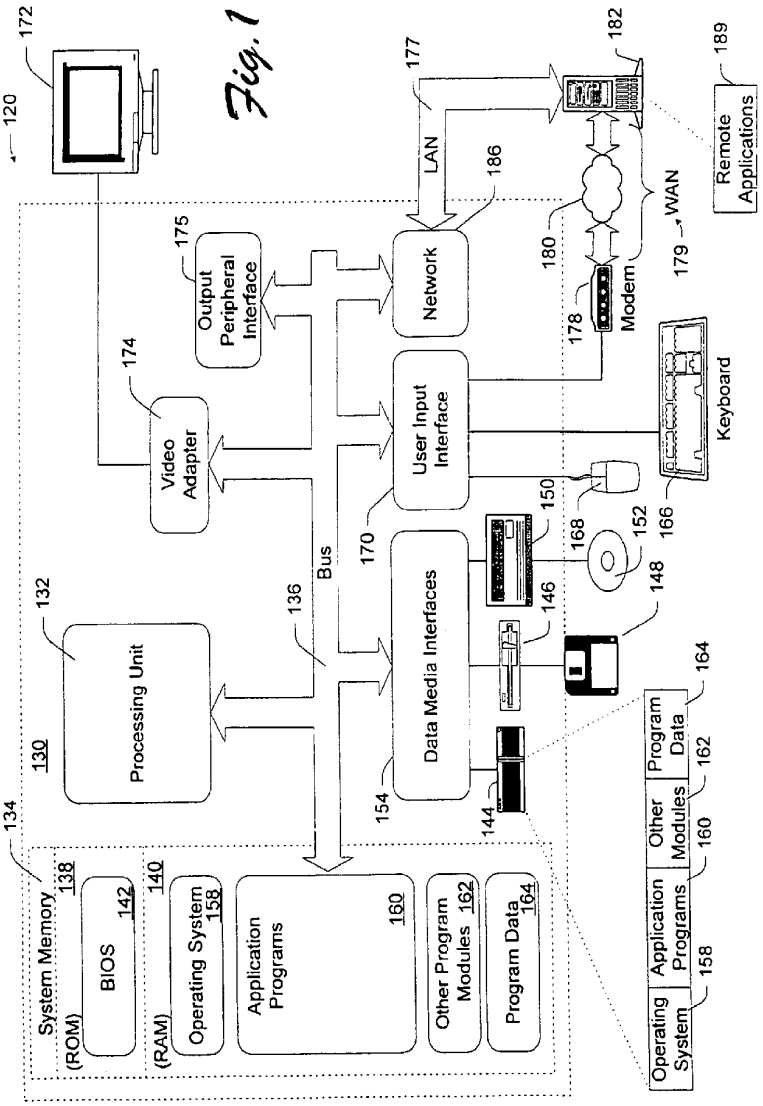
17. The method of claim 15 wherein the plural surrounding portions are plural surrounding macroblocks.

2003204477 29 Jan 2009

18. The method of claim 1 wherein, for each of the at least one direct mode macroblock in the B-picture, the spatial motion vector prediction uses motion information of surrounding macroblocks within the B-picture in motion vector prediction, or the temporal motion vector prediction uses motion information of a collocated macroblock in a P-picture.

19. The method of claim 10 wherein, for each of the at least one direct mode macroblock in the B-picture, the spatial motion vector prediction uses motion information of surrounding macroblocks within the B-picture in motion vector prediction, or the temporal motion vector prediction uses motion information of a collocated macroblock in a P-picture.

20. A method substantially as hereinbefore described with reference to the accompanying drawings.


120
172
Fig. 1
177
182
189
LAN
186
Remote Applications
175
Output Peripheral Interface
Network
180
WAN
179
174
Video Adapter
User Input Interface
178
Modem
Keyboard
Bus
170
150
168
152
166
132
Processing Unit
136
Data Media Interfaces
146
148
130
154
144
134
System Memory
138
(ROM)
BIOS
142
140
(RAM)
Operating System
158
Application Programs
160
Other Program Modules
162
Program Data
164
Operating System
158
Application Programs
160
Other Modules
162
Program Data
164

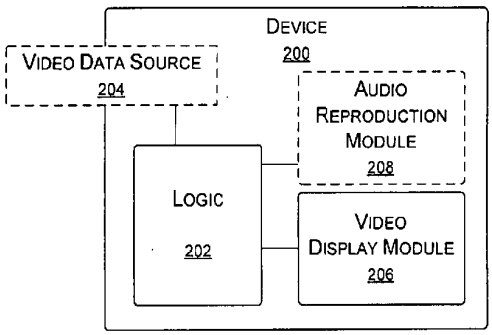

DEVICE
200
VIDEO DATA SOURCE
204
AUDIO
REPRODUCTION
MODULE
208
LOGIC
202
VIDEO
DISPLAY MODULE
206

Fig. 2

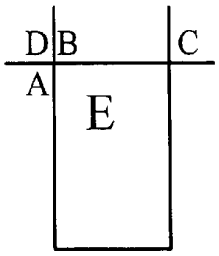

D
B
C
A
E

Fig. 3

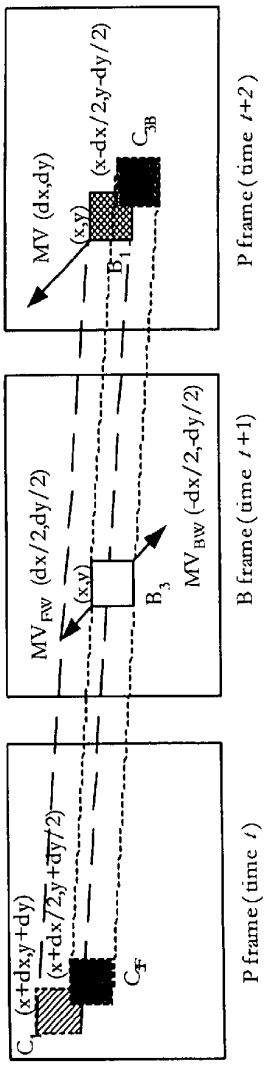

P frame (time t+2)
MV (dx,dy)
(x,y)
(x-dx/2,y-dy/2)
C_{3B}
B_1
B frame (time t+1)
MV_{FW} (dx/2,dy/2)
(x,y)
B_3
MV_{BW} (-dx/2,-dy/2)
P frame (time t)
C_1
(x+dx,y+dy)
(x+dx/2,y+dy/2)
C_F

Fig. 4

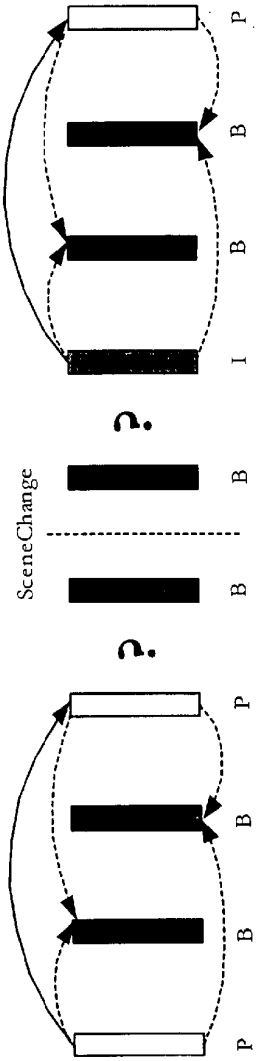

P
B
B
I
B
B
P
SceneChange
P
B
B
P
B
B
P

Fig. 5

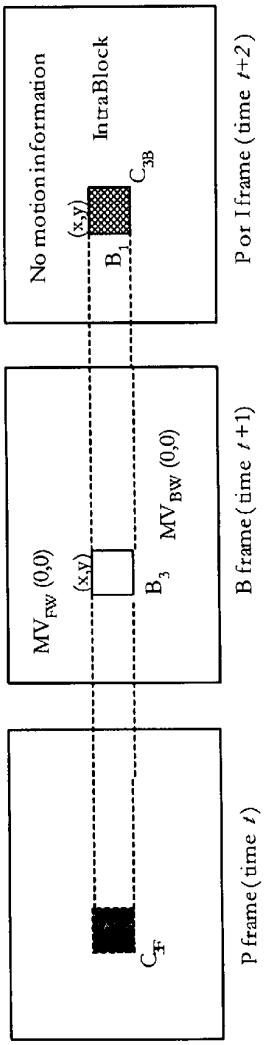

No motion information
Intra Block
(x,y)
B_1
C_{3B}
P or I frame (time t+2)
MV_{FW} (0,0)
(x,y)
B_3
MV_{BW} (0,0)
B frame (time t+1)
C_F
P frame (time t)

Fig. 6

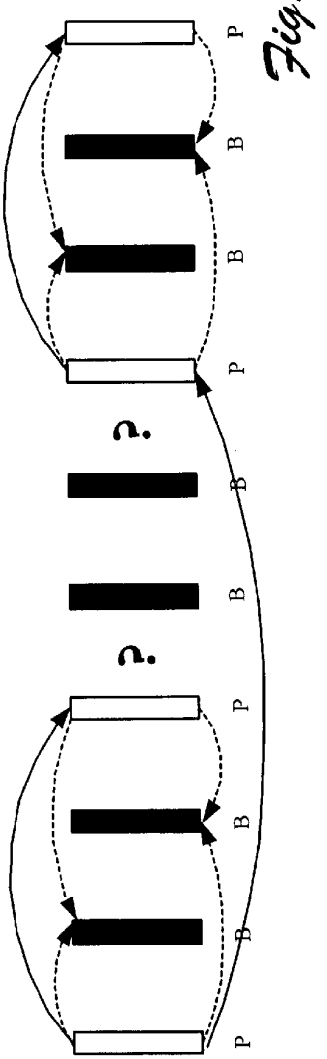

P
B
B
P
B
B
P
B
B
P
B
B
P
B
B
P

Fig. 7

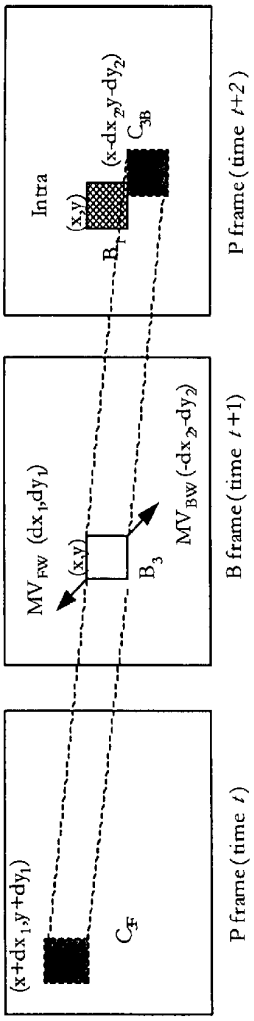

Intra
(x,y)
B1
(x-dx2,y-dy2)
CBB
P frame (time t+2)
MVFW (dx1,dy1)
(x,y)
B3
MVBW (-dx2,-dy2)
B frame (time t+1)
(x+dx1,y+dy1)
CF
P frame (time t)

Fig. 8

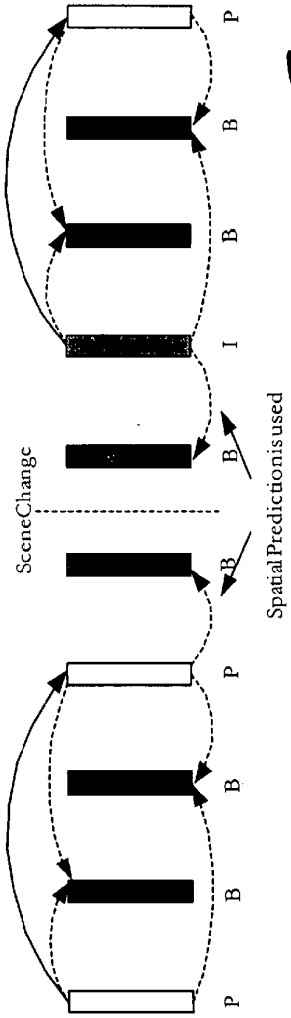

P
B
B
P
B
B
I
B
B
P
B
B
P
Scene Change
Spatial Prediction is used

Fig. 9

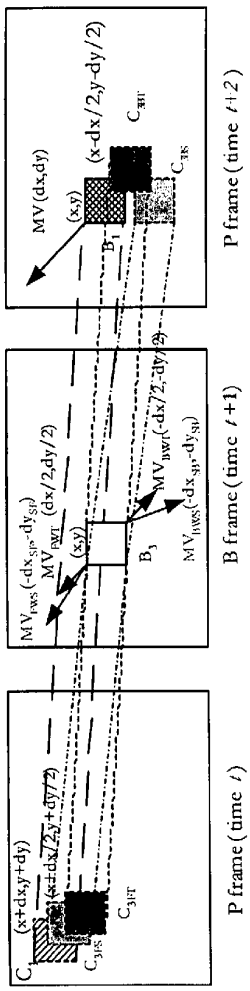

P frame (time t)
B frame (time t+1)
P frame (time t+2)
MV(dx,dy)
(x,y)
(x-dx/2,y-dy/2)
(x+dx,y+dy)
(x+dx/2,y+dy/2)

Fig. 10

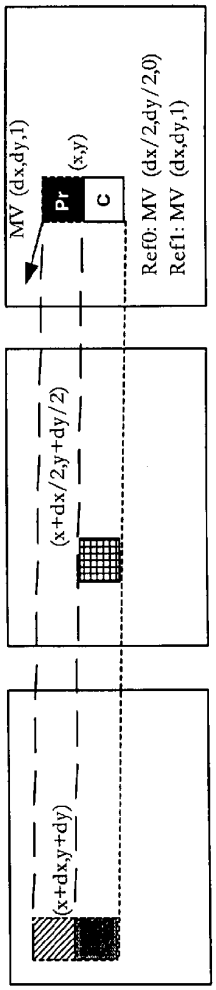

Ref1 (time t)
Ref0 (time t+1)
Current (time t+2)
MV (dx,dy,1)
Pr
C
(x,y)
(x+dx,y+dy)
(x+dx/2,y+dy/2)
Ref0: MV (dx/2,dy/2,0)
Ref1: MV (dx,dy,1)

Fig. 11

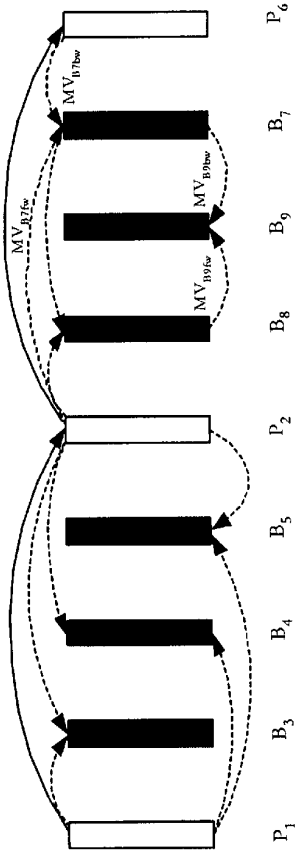

P1
B3
B4
B5
P2
B8
B9
B7
P6
MV_B7fw
MV_B7bw
MV_B9fw
MV_B9bw

Fig. 12

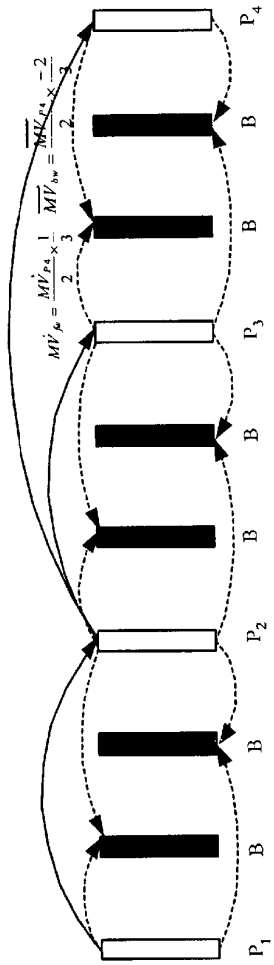

P1
B
B
P2
B
B
P3
B
B
P4
MV̇P4
MV̇fw = MV̇P4 × 1/3
MV̇bw = MV̇P4 × −2/3

Fig. 13

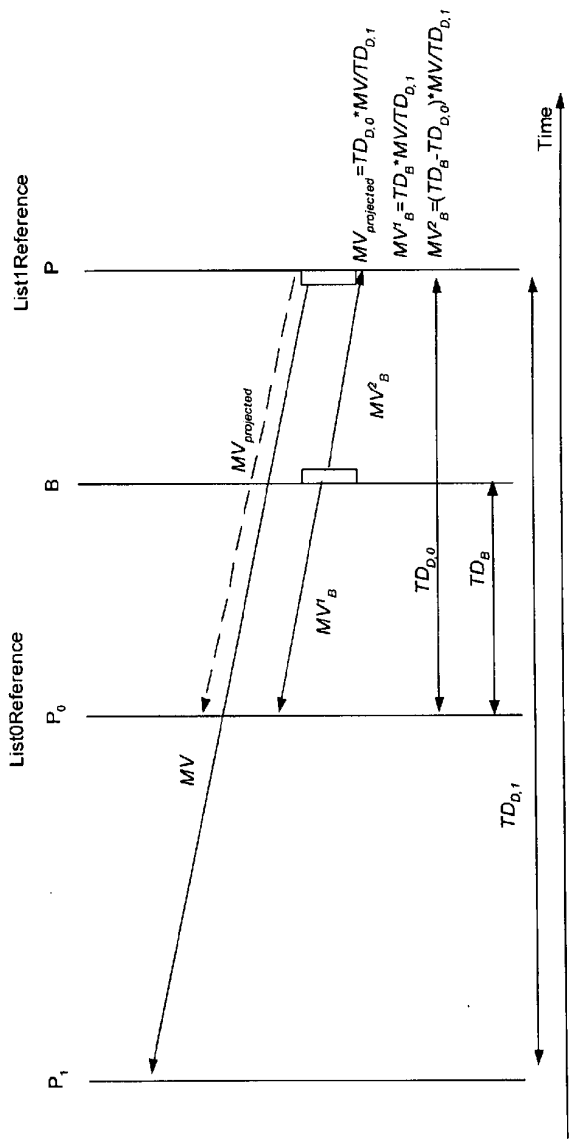

List1Reference
P
B
List0Reference
P_0
P_1
Time
$MV_{projected}=TD_{D,0}*MV/TD_{D,1}$
$MV^1_B=TD_B*MV/TD_{D,1}$
$MV^2_B=(TD_B-TD_{D,0})*MV/TD_{D,1}$
$MV_{projected}$
MV^2_B
MV^1_B
$TD_{D,0}$
TD_B
MV
$TD_{D,1}$

Fig. 14

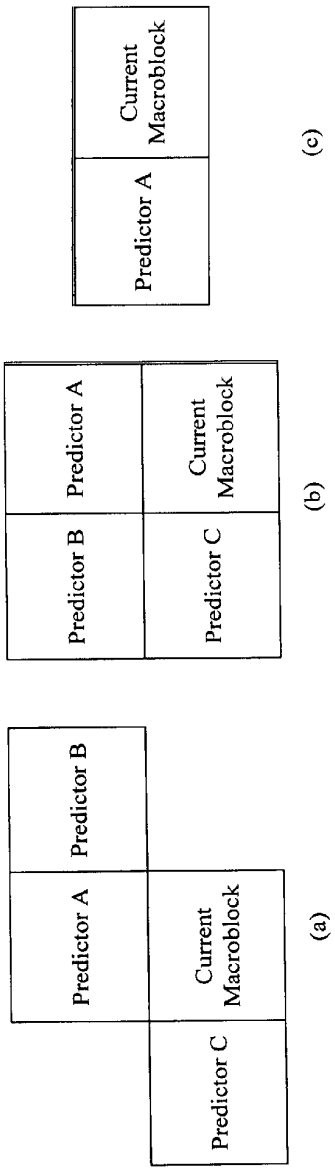

Predictor B
Predictor A
Current Macroblock
Predictor C
(a)
Predictor A
Predictor B
Current Macroblock
Predictor C
(b)
Predictor A
Current Macroblock
(c)

Fig. 15

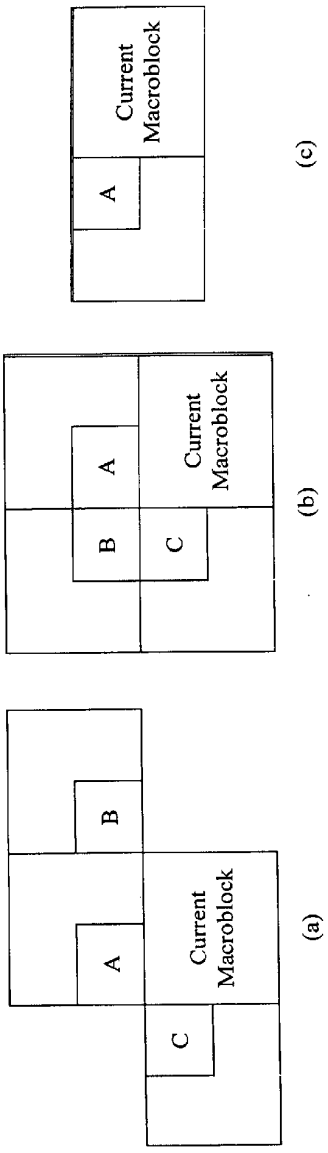

Current Macroblock
A
(c)
Current Macroblock
A
B
C
(b)
Current Macroblock
A
B
C
(a)

Fig. 16

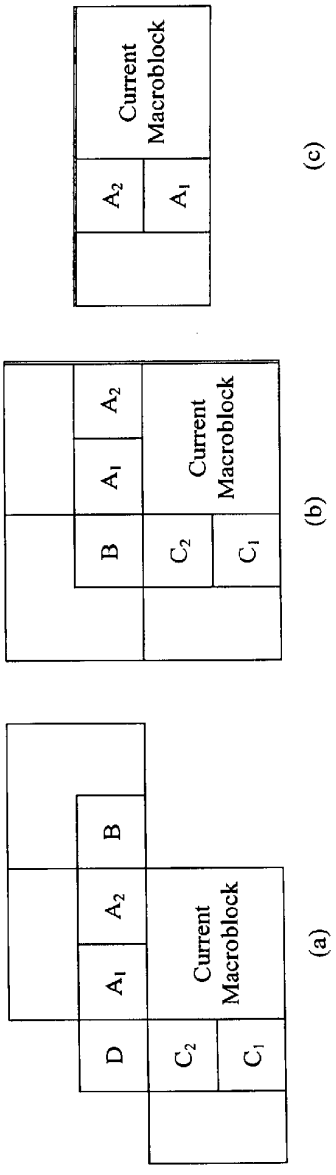

D
B
A_2
A_1
Current Macroblock
C_2
C_1
(a)
B
A_2
A_1
Current Macroblock
C_2
C_1
(b)
A_2
A_1
Current Macroblock
(c)

Fig. 17