

(51) International Patent Classification:
H04N 7/01 (2006.01)(21) International Application Number:
PCT/IB2009/055097(22) International Filing Date:
22 September 2009 (22.09.2009)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
PCT/CN2008/072447
22 September 2008 (22.09.2008) CN(71) Applicant (for all designated States except US):
FRANCE TELECOM [FR/FR]; 6, Place d'Alleray,
F-75015 Paris (FR).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **WANG, Ronggang**
[CN/CN]; 10F, South Twr, Raycom InfoTech Park C, 2
Science Institute South Rd, Haidian District, Beijing
100080 (CN). **ZHANG, Yongbing** [CN/CN]; 10/F South
Tower Raycom Info Tech ParkC, 2 Science Institute-
South Rd, Haidian District, Beijing 100080 (CN). **PA-
TEUX, Stéphane** [FR/FR]; 1 Allée Mathurin Hardy De
La Largère, F-35700 Rennes (FR).(74) Agent: **FRANCE**
TELECOM/FTR&D/PIV/BREVETS; BORNIER Ray-
nald, 38-40 Rue du Général Leclerc, F-92794 Issy
Moulineaux Cedex 9 (FR).(81) Designated States (unless otherwise indicated, for every
kind of national protection available): AE, AG, AL, AM,
AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ,
CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO,
DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT,
HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP,
KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD,
ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI,
NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD,
SE, SG, SK, SL, SM, ST, SV, SY, TJ, TM, TN, TR, TT,
TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.(84) Designated States (unless otherwise indicated, for every
kind of regional protection available): ARIPO (BW, GH,
GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM,
ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ,
TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE,
ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV,
MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM,
TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW,
ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

[Continued on next page]

(54) Title: FRAME RATE UP CONVERSION METHOD AND SYSTEM

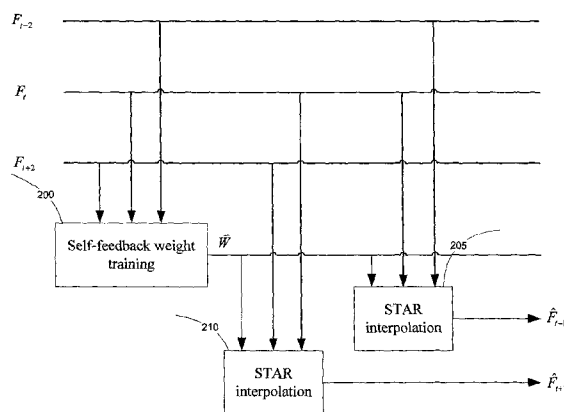


Figure 2

(57) **Abstract:** A method for adding frames in a digital video flow of frames, the added frames comprising a first predicted frame added between a first and a second frames of the digital video flow, and a second predicted frame added between a second and a third frames of said digital video flow, the method further comprising the acts of: a) initializing an intermediate estimate of the first and second predicted frames, and for each intermediate estimate of the first and second predicted frames: b1) defining a new intermediate estimate of the first predicted frame using a first interpolation function applied to the first and second frames, b2) defining a new intermediate estimate of the second predicted frame using a second interpolation function applied to the second and third frames, b3) defining a prediction of the second frame with a third interpolation function applied to the intermediate estimate of the first and second predicted frames, b4) optimizing the interpolation functions while minimizing the distortion between the new intermediate estimate and the intermediate estimate of the predicted frames and between the second frame and its prediction, b5) computing the intermediate estimates of the first and second predicted frames using the optimized first and second interpolation functions, the first and second predicted frames being set to the intermediate estimates of the first and second predicted frames when the distortion is smaller than a given threshold or a given number of intermediate estimates of the first and second predicted frames has been computed.



-
- | | | |
|---|-------------------|--|
| — as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii)) | Published: | — with international search report (Art. 21(3)) |
| — as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii)) | — | before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h)) |
| — of inventorship (Rule 4.17(iv)) | | |

FRAME RATE UP CONVERSION METHOD AND SYSTEM

Field of the Invention

The present invention relates in general to image processing and more specifically to Frame Rate Up-Conversion (FRUC) image processing.

Background of the Invention

In image processing, Frame Rate Up-Conversion (FRUC) is a widely investigated technique to up-convert the temporal resolution of a media sequence (e.g. a video), i.e. to increase the number of frames (or pictures or images) in a media sequence in order to increase its resolution. FRUC has many applications, among which the most practical one is to enhance the visual quality of low bit rate coded video. In low bit rate coding applications, it is possible to reduce the frame rate to the half or even lower ratio of the original frame rate, and then only encode the low rate frames with better visual quality. When decoding the media, FRUC may be performed to restore the loss of temporal resolution with higher visual quality. One of the most common cases is the double up-conversion of the frame rate, an example of which is illustrated on Figure 1, where for instance, the original frame rate is 15 fps (frames per second), and the up-converted frame rate is 30 fps. The added frames are interpolated using the redundancies between successive frames along the temporal axis.

Besides the low bit rate coding, FRUC is also applicable for instance to Phase Alternating Line-National Television System Committee (PAL-NTSC) conversion and complex video editing. In addition, FRUC may be of great help for slow-motion playback and the rate allocation policy of scalable video coding schemes.

Numerous FRUC algorithms have been developed to convert frame rates. A simple and straightforward FRUC algorithm consists in combining adjacent video frames without taking object motion into account. The object motion is the change of a given object, captured from a scene, from one frame to another (i.e. between adjacent frames). Examples of such a simple FRUC method are e.g. frame repetition or frame averaging.

However, such methods work well only if there is little or no motion between successive or adjacent frames. As the intensity of motion increases, the frame repetition method will cause jerkiness and the resulting image will look choppy and/or not smooth. When applying a frame averaging method, images will look
5 blurry when motion occurs, and sometimes ghost artifacts may also be perceived.

As opposed to frame repetition and frame averaging, another kind of FRUC method, called a MC (Motion Compensation)-FRUC, is based on Motion Estimation (ME) and performs frame interpolation along the motion trajectory of an object from frame to frame to achieve better visual quality. This method uses motion vectors, i.e.
10 vectors used for inter-prediction that provides an offset from the coordinates in the decoded picture to the coordinates in a reference picture. Since motion plays a significant role in MC-FRUC, many algorithms have been developed to derive accurate motion vectors. For example, the Block-Matching Algorithm (BMA) detailed in "A method for motion adaptive frame rate up-conversion" R. Castagno, P.
15 Haavisto, and G. Ramponi, *IEEE Trans. on Circuits and System Video Technology*, vol. 6, No. 5, pp. 436-446, Oct. 1996 and "Adaptive MC interpolation for frame rate up-conversion" S.-H. Lee, Y.-C. Shin, S.-J. Yang, H.-H. Moon and R.-H. Park, *IEEE Trans. on Consumer Electron.*, vol. 48, No. 3, pp. 444-450, Aug. 2002, has a broad application in MC-FRUC. Since motion vectors derived by BMA are often not
20 accurate enough, several approaches for more faithful motion estimation, described in "New frame rate up-conversion using bi-directional motion estimation", B. T. Choi, S. H. Lee, and S. J. Ko, *IEEE Trans. on Consumer Electron.*, vol. 46, No. 3, pp. 603-609, Aug. 2000 and "Motion compensated frame interpolation based on H.264 decoder", Z. Gan, L. Qi, and X. Zhu, *Electronics Letters.*, vol. 43, No. 2, pp. 96-98,
25 Jan. 2007, have also been proposed in recent works. One of these FRUC algorithms "New frame rate up-conversion using bi-directional motion estimation", B. T. Choi, S. H. Lee, and S. J. Ko, *IEEE Trans. on Consumer Electron.*, vol. 46, No. 3, pp. 603-609, Aug. 2000, uses a bi-directional ME to derive more faithful motion vectors. In addition, a hierarchical MC technique was proposed to achieve better
30 visual quality in "Hierarchical motion compensated frame rate up-conversion based

on the Gaussian/Laplacian pyramid”, G. I. Lee, B. W. Jeon, R. H. Park, and S. H. Lee, *Proc. IEEE Int. Conf. Consumer Electronics*, 2003, pp. 350-351. Both methods achieve better performance than BMA does. However, they are all based on the assumption of a translational motion with constant velocity, which is not always true.

5 Furthermore, constant acceleration was exploited to derive more reliable motion trajectories. However, the assumption of a constant acceleration also does not always hold true for all the regions of an image, e.g. for the regions of non-rigid objects (i.e. for instance clouds and waves which change their forms/shape between images).

10 As opposed to its use in video coding where it is used with actual frames, Motion Estimation may be performed for FRUC with the absence of actual frames, i.e. it allows creating some frames between and from existing ones. However, as the derived motion trajectory may not be consistent sometimes, a median filter is usually applied to the motion vector field (i.e. the field of motion vectors of all the
15 blocks) after performing ME, in which the motion vector field needs to go through the median filter operation in e.g. a 3x3 window (or current block and its 8 neighboring blocks which are located in a 3x3 window) before MC. However, for the areas of small objects, irregular shaped objects, and object boundaries, fixed size block MC usually does not work well. To avoid this situation, variable size block MC
20 was proposed in object boundaries to reconstruct edge information with a higher quality. Furthermore, a method called Overlapped Block MC (OBMC) may be applied to suppress the blocking artifacts, which are usually observed when a block has a significant different motion vector compared with its neighboring blocks. However, OBMC may sometimes over-smooth the edges of the image and degrade
25 the image quality. To reduce the over-smoothing effect of OBMC, an Adaptive OBMC (AOBMC) method was proposed, in which the coefficients of the OBMC are adjusted based on the reliability of neighboring motion vectors. However, AOBMC still has poor ability to represent some complex motions, such as zooming, rotation, and local deformation.

30 Today there is a need for a FRUC solution that can be easily implemented on the existing communication infrastructures, overcoming the drawbacks of the prior art.

Summary of Invention

It is an object of the present system to overcome disadvantages and/or make improvement over the prior art.

5 To that extend, the invention proposes a method A method for adding frames in a digital video flow of frames, the added frames comprising a first predicted frame added between a first and a second frames of the digital video flow, and a second predicted frame added between a second and a third frames of said digital video flow, the method further comprising the acts of:

10 a) initializing an intermediate estimate of the first and second predicted frames,

and for each intermediate estimate of the first and second predicted frames:

b1) defining a new intermediate estimate of the first predicted frame using a first interpolation function applied to the first and second frames,

15 b2) defining a new intermediate estimate of the second predicted frame using a second interpolation function applied to the second and third frames,

b3) defining a prediction of the second frame with a third interpolation function applied to the intermediate estimate of the first and second predicted frames,

20 b4) optimizing the interpolation functions while minimizing the distortion between the new intermediate estimate and the intermediate estimate of the predicted frames and between the second frame and its prediction,

b5) computing the intermediate estimates of the first and second predicted frames using the optimized first and second interpolation functions,

25 the first and second predicted frames being set to the intermediate estimates of the first and second predicted frames when the distortion is smaller than a given threshold or a given number of intermediate estimates of the first and second predicted frames has been computed.

The invention also relates to a system for adding frames in a digital video flow of frames, the added frames comprising a first predicted frame added between a first and a second frames of the digital video flow, and a second predicted frame added between a second and a third frames of said digital video flow, said system comprising:

- an emitting device operable to send said digital video flow,

- a receiving device operable to:

a) initialize an intermediate estimate of the first and second predicted frames, and for each intermediate estimate of the first and second predicted frames:

10 b1) define a new intermediate estimate of the first predicted frame using a first interpolation function applied to the first and second frames,

b2) define a new intermediate estimate of the second predicted frame using a second interpolation function applied to the second and third frames,

15 b3) define a prediction of the second frame with a third interpolation function applied to the intermediate estimate of the first and second predicted frames,

b4) optimize the interpolation functions while minimizing the distortion between the new intermediate estimate and the intermediate estimate of the predicted frames and between the second frame and its prediction,

20 b5) compute the intermediate estimates of the first and second predicted frames using the optimized first and second interpolation functions,

the first and second predicted frames being set to the intermediate estimates of the first and second predicted frames when the distortion is smaller than a given threshold or a given number of intermediate estimates of the first and second predicted frames has been computed;

25 - a network operable to communicate between the emitting device and the receiving device.

The invention also relates to an interpolating device for adding frames in a digital video flow of frames, the added frames comprising a first predicted frame added between a first and a second frames of the digital video flow, and a second

30

predicted frame added between a second and a third frames of the digital video flow, said device being operable to receive said digital video flow of frames, the device being further operable to:

- a) initialize an intermediate estimate of the first and second predicted frames,
 - 5 and for each intermediate estimate of the first and second predicted frames:
 - b1) define a new intermediate estimate of the first predicted frame using a first interpolation function applied to the first and second frames,
 - b2) define a new intermediate estimate of the second predicted frame using a second interpolation function applied to the second and third frames,
 - 10 b3) define a prediction of the second frame with a third interpolation function applied to the intermediate estimate of the first and second predicted frames,
 - b4) optimize the interpolation functions while minimizing the distortion between the new intermediate estimate and the intermediate estimate of the predicted frames and between the second frame and its prediction,
 - 15 b5) compute the intermediate estimates of the first and second predicted frames using the optimized first and second interpolation functions,
- the first and second predicted frames being set to the intermediate estimates of the first and second predicted frames when the distortion is smaller than a given threshold or a given number of intermediate estimates of the first and second predicted frames has been computed.
- 20

The invention also relates to a computer program providing computer executable instructions stored on a computer readable medium, which when loaded on to a data processor causes the data processor to perform a method for adding frames in a digital video flow of frames according to claims 1 to 3.

25

The method according to the invention proposes a model for improving the quality of interpolated frames in FRUC. The present method could be described as a Spatiotemporal Auto-Regressive (STAR) model.

Indeed, the proposed method exploits the correlations or similarities (of pixels) between adjacent frames (i.e. in the temporal domain) and within a same frame (i.e. spatial domain).

5 The method according to the invention is able to adaptively exploit the redundancies both in the spatial and temporal domains by tuning the interpolation functions (i.e. its interpolation weights) in order to optimize them, making thus interpolated frames more reliable. Indeed, the STAR model allows learning motion information between successive frames by adjusting the interpolation weights adaptively according to the characteristics of the pixels within a local spatiotemporal
10 neighborhood.

Another advantage of the method according to the invention is that it achieves sub-pixel (i.e. pixels located between the space of integer pixels) accuracy with spatially varying interpolation coefficients or weights. As traditional FRUC algorithms use the same interpolation coefficients for all the frames of a sequence,
15 it is not possible to consider non-stationary statistical properties of video signals (i.e. frames or areas with different motion intensity, for example, the boundary of motion objects, where one part of the object is stationary and the other part is moving, or the boundary of foreground and background regions, or the covered and uncovered regions between successive frames). In contrast, the STAR model is able to tune its
20 interpolation coefficients adaptively to match non-stationary statistical properties of video signals.

Brief Description of the Drawings

Embodiments of the present invention will now be described solely by way of
25 example and only with reference to the accompanying drawings, where like parts are provided with corresponding reference numerals, and in which:

Figure 1 schematically illustrates a simple example of Frame Rate Up-Conversion (FRUC);

Figure 2 schematically illustrates a system according to an embodiment of
30 the present invention;

Figure 3 schematically illustrates an example of pixels interpolation according to an embodiment of the present invention;

Figure 4 schematically illustrates the method according to an embodiment of the present invention;

Figure 5 schematically illustrates the self-feedback weight training method according to an embodiment of the present invention;

5 Figure 6 schematically illustrates an example of pixels interpolation according to an embodiment of the present invention;

Figure 7 schematically illustrates an example of Motion Compensation Interpolation (MCI) method according to an embodiment of the present invention

10 Figure 8 schematically illustrates a simulation according to an embodiment of the present invention.

Description of Preferred Embodiments

15 The method according to the invention proposes a model for improving the quality of interpolated frames in FRUC, and more specifically a Spatiotemporal Auto-Regressive (STAR) model.

In the STAR model, each pixel is approximated as a weighted linear combination or interpolation (through interpolation functions) of pixels within a spatial neighborhood in the current frame as well as a weighted linear combination of pixels within a temporal neighborhood in the previous and following frames.

20 Due to the absence of some actual pixels in the missing frames and in order to improve the accuracy of the model, the method according to the invention proposes an iterative self-feedback weight training algorithm to optimize the interpolation weights (i.e. the coefficients) of the linear combinations. An initialization step allows interpolating at least one pixel in a given missing frame to
25 be interpolated, hereafter called first and second predicted frames. Then, in each iteration, pixels in said given missing frame are interpolated using pixels in the previous and following frames as well as available pixels in the given missing frame. Actual original low rate frames are also predicted (i.e. estimated) using the same interpolation weights in the same iteration. Finally, after a pre-set number of
30 iterations or when the distortion between pixels from one iteration to another is minimized (i.e. under a given threshold), then the optimal interpolation weights are

derived. The distortion is calculated between the pixels approximated using the interpolation weights in the current iteration and those in the previous iteration as well as the actual pixels in the original low rate frame.

5 According to an exemplary embodiment of the present invention, the STAR model is based on a spatiotemporal scheme, which is depicted in Figure 3.

Along the temporal axis, model samples are taken at five successive frames , where F_{t-2} , F_t , and F_{t+2} denote the successive original low rate frames, respectively called first, second and third frames, $\hat{F}_{t-1}$ and $\hat{F}_{t+1}$ denote the interpolated or added
10 frames at time instance $t-1$ and $t+1$, respectively referred to as first and second predicted frames.

The pixel value to be interpolated at spatial location (m, n) of a current frame, the current frame being either $\hat{F}_{t-1}$ or $\hat{F}_{t+1}$, depends on pixel samples within a temporal neighbourhood taken in previous and following frames as well as pixels
15 available within a spatial neighbourhood of said current frame. When the current frame is $\hat{F}_{t-1}$, the previous and following frames are respectively the first frame F_{t-2} and the second frame F_t . When the current frames is $\hat{F}_{t+1}$, the previous and following frames are respectively the second frame F_t and the third frame F_{t+2} .

The temporal neighbourhood or N_T can be defined as a region taken in the
20 previous or following frames to the current frame. The spatial neighbourhood or N_S can be defined as a region taken in the current frame where pixels are already available as they have been previously defined.

The size of the temporal neighbourhood N_T in the previous and following frames may defined as a $(2L+1) \times (2L+1)$ square bounded by $(m \pm L, n \pm L)$. To be
25 consistent with N_T , the size of the spatial neighborhood N_S may be chosen as being of the same size as N_T or smaller.

Parameter L is referred to as the spatiotemporal order of STAR model. Figure 3 shows an exemplary illustration of the STAR model using a spatiotemporal order $L=1$.

30 The method according to the invention may be performed on a block within the frame, block defined as the supporting region S of the STAR model. It may correspond for example to a block of pixels (for instance 16x16 or 32x32 pixel

blocks) or to the whole picture. As seen hereafter within a block, the interpolation weights are identical per iteration and over N_T or N_S . In other words, for every (m,n) in the supporting region, the contribution of the pixels of N_T or N_S in the first to third frames only depends upon their distance to (m,n) .

5

Figure 2 describes an illustrative embodiment of the system according to the invention.

A self-feedback weight training unit 200 allows deriving the optimal interpolation weights by iteration. Each pixel is approximated as a weighted linear combination of pixels (through interpolation functions) within a spatial neighborhood in the current frame as well as a weighted linear combination of pixels within a temporal neighborhood in the previous and following frames. At the end of each iteration, optimal interpolation weights are derived. The self-feedback weight training process is later described in detail hereunder with reference to Figure 5.

After getting the optimal interpolation weights, a first unit 205 allows interpolating the missing frame at time instance $t-1$ and a second unit 210 allows interpolating the missing frame at time instance $t+1$.

Figure 4 describes an illustrative embodiment of the method according to the invention.

According to the spatial resolution of the input sequences (i.e. frame resolution), the size of the supporting region is first determined in an act 400. The supporting region is the area of a frame on which the method according to the invention may be performed.

According to an illustrative embodiment of the present invention, the determination process may be as follows: for example, if the resolution is QCIF (Quarter Common Intermediate Format), then the size of the supporting region may be set to be 16x16, else if the resolution is larger than QCIF, the size of the supporting region is set to be 32x32. Then, as long as the frames interpolation process is not completed then (act 410) the spatiotemporal order L of the STAR model is computed in an act 420 as

$$L = \max_{i \in S} \left\{ \text{abs}(\lfloor mvx_i \rfloor), \text{abs}(\lfloor mvy_i \rfloor) \right\} + 1, \quad (1)$$

where S represents the supporting region, $\lfloor \dots \rfloor$ is the floor function, which maps the motion vector mv_i of the i th block (i.e. the block used to interpolate the to-be-interpolated frame by the MCI method, in other words it means the blocks used to generate F_{t-1}^0 and F_{t+1}^0 as shown in equation (7)) in S to the full-pixel position.

5 After the supporting region and the spatiotemporal order have been determined, the self-feedback weight training process is performed in an act 430 to derive the optimal interpolation weights. Finally, the missing frames $\hat{F}_{t-1}$ and $\hat{F}_{t+1}$ are interpolated by STAR interpolation using the optimal weights in an act 440.

10

Figure 5 describes an illustrative embodiment of the self-feedback weighting training process of act 430 according to the invention. This training process is an iterative process. In the hereafter description, the supporting region S or block is assumed to be the whole frame. The present teaching may be easily transposed by the man skilled in the art to a supporting region taken in a frame.

Along the temporal axis, the three successive first, second and third frames of the original low rate frames (respectively F_{t-2} , F_t and F_{t+2}) are selected in an act 500.

20 $\hat{F}_{t-1}^{i+1}$ and $\hat{F}_{t+1}^{i+1}$ denote intermediate values or estimates for the first and second predicted frames for the i^{th} iteration. i represents the iteration counter and $\bar{W}^i$ represents the optimal weight vector after the i^{th} iteration.

Before iterations begin, a preliminary act 510 of initialization is performed. The act 510 allows initializing $\hat{F}_{t-1}^0$ and $\hat{F}_{t+1}^0$. Various solutions may be implemented to initialize $\hat{F}_{t-1}^0$ and $\hat{F}_{t+1}^0$ (for example, initial values may be pre-set or pre-defined). In this illustrative embodiment according to the present invention, MCI (Motion Compensation Interpolation) is used. MCI is further described hereunder with reference to Figure 7. The iteration counter i is also initialized to 0.

30 The iterative process is then started as follows. In an additional act 520, the iteration counter i is compared to a pre-defined maximum value i_{Max} . If the

maximum value has not been reached, an optimal weight vector $\bar{W}^i$ is further derived in an act 530. This weight vector $\bar{W}^i$ allows computing the interpolated values of frames $\hat{F}_{t-1}^{i+1}$ and $\hat{F}_{t+1}^{i+1}$, respectively according to equations (2) and (3) here below:

5

$$\begin{aligned} \hat{f}_{t-1}^{i+1}(m, n) = & \sum_{(k,l) \in N_T} f_{t-2}(m+k, n+l) \times w_f^i(k, l) \\ & + \sum_{(u,v) \in N_T} f_t(m+u, n+v) \times w_b^i(u, v) \\ & + \sum_{(e,f) \in N_S} \hat{f}_{t-1}^{i+1}(m+e, n+f) \times w_s^i(e, f) \end{aligned} \quad (2)$$

$$\begin{aligned} \hat{f}_{t+1}^{i+1}(m, n) = & \sum_{(k,l) \in N_T} f_t(m+k, n+l) \times w_f^i(k, l) \\ & + \sum_{(u,v) \in N_T} f_{t+2}(m+u, n+v) \times w_b^i(u, v) \\ & + \sum_{(e,f) \in N_S} \hat{f}_{t+1}^{i+1}(m+e, n+f) \times w_s^i(e, f) \end{aligned} \quad (3)$$

10 wherein any $f(i,j)$ denotes the pixel value at (i,j) for frame F and (m,n) a pixel on the chosen supporting region of a frame.

In an additional embodiment of the method according to the invention, the first and the second predicted frames at iteration i ($\hat{F}_{t-1}^{i+1}$ and $\hat{F}_{t+1}^{i+1}$) may be used to generate a prediction $\hat{F}_t^{i+1}$ of the second frame F_t according to equation

15 (4) here below:

$$\begin{aligned}
\hat{f}_t^{i+1}(m, n) = & \sum_{(k,l) \in N_T} \hat{f}_{t-1}^i(m+k, n+l) \times w_f^i(k, l) \\
& + \sum_{(u,v) \in N_T} \hat{f}_{t+1}^i(m+u, n+v) \times w_b^i(u, v) \\
& + \sum_{(e,f) \in N_S} \hat{f}_t^{i+1}(m+e, n+f) \times w_s^i(e, f)
\end{aligned} \tag{4}.$$

In equations (2) to (4), $w_f^i(k, l)$, $w_b^i(u, v)$ and $w_s^i(e, f)$ represent the forward, backward, and the spatial weight components of the weight vector $\bar{W}^i$ corresponding to the previous, following, and current frames, respectively.

5 $w_f^i(k, l)$, $w_b^i(u, v)$ are defined over N_T as they correspond to the weights for the previous and following frames, while $w_s^i(e, f)$ are defined over N_S as they correspond to the weights for the current frame in the interpolation functions of equations (2) to (4). These equations (2) to (4) can be seen as a weighted sum or interpolation functions applied to frames from the sequence of original low

10 rate frames. In the present illustration, the interpolation functions are chosen as identical functions for equations (2), (3) and (4). More generally 3 interpolations functions may be chosen for (2), (3), and (4) respectively.

Furthermore, in equations (2), (3) and (4), the interpolation function is applied to the already defined pixels. This indeed corresponds to the spatial

15 neighborhood N_S as illustrated in Figure 3.

For instance, as $\hat{F}_{t-1}^{i+1}$ is defined through (2), N_S increases in size as more and more pixels are defined in $\hat{F}_{t-1}^{i+1}$, consequently the contribution of $\hat{F}_{t-1}^{i+1}$ pixels in its own definition increases. In an additional embodiment of the present method, the size of N_S is limited to N_T , as mentioned before to be consistent with the

20 size of the temporal neighborhood.

In a further embodiment, for instance if the pixels in a supporting region S are defined line by line or column by column, always in the same direction (left to right, right to left, top to bottom, ...), in the current spatial neighborhood, only $\frac{1}{2}[(2L+1) \times (2L+1) - 1]$ pixels have been defined through equations (2), (3)

and (4). Therefore, N_s may be limited in size to the $\frac{1}{2}[(2L+1) \times (2L+1) - 1]$ defined pixels.

In equation (4), one may see that $\hat{F}_t^{i+1}$ is a weighted combination of the intermediate values of the first and second predicted frames from the previous iteration.

The weight components are variables that may be optimized through equation (5):

$$\bar{W}^i = \arg \min_{\bar{W}^i} \left\{ \left\| \hat{F}_{t-1}^i - \hat{F}_{t-1}^{i+1} \right\|^2 + \left\| \hat{F}_{t+1}^i - \hat{F}_{t+1}^{i+1} \right\|^2 + \left\| F_t - \hat{F}_t^{i+1} \right\|^2 \right\} \quad (5)$$

wherein the **argmin** function give the optimal weight vector $\bar{W}^i$ minimizing the sum or distortion $\left\| \hat{F}_{t-1}^i - \hat{F}_{t-1}^{i+1} \right\|^2 + \left\| \hat{F}_{t+1}^i - \hat{F}_{t+1}^{i+1} \right\|^2 + \left\| F_t - \hat{F}_t^{i+1} \right\|^2$.

Equation (5) allows optimizing the interpolation functions while minimizing the distortion between the new intermediate estimate and the intermediate estimate of the predicted frames (elements $\left\| \hat{F}_{t-1}^i - \hat{F}_{t-1}^{i+1} \right\|^2$ and $\left\| \hat{F}_{t+1}^i - \hat{F}_{t+1}^{i+1} \right\|^2$ respectively of (5)) and between the second frame and its prediction (element $\left\| F_t - \hat{F}_t^{i+1} \right\|^2$ of (5)).

The resulting optimal weights $\bar{W}^i_{\text{optimal}}$ are used in a further act 540 to generate the intermediate values of the first and second predicted frames, respectively $\hat{F}_{t-1}^{i+1}$ and $\hat{F}_{t+1}^{i+1}$, as well as the prediction of the second frame $\hat{F}_t^{i+1}$, for iteration i .

To decide whether the exemplary embodiment of the present method will be terminated in the current iteration i , a distortion taking into account the frames $\hat{F}_{t-1}^{i+1}$, $\hat{F}_{t+1}^{i+1}$ and $\hat{F}_t^{i+1}$ calculated in act 540 for iteration i , as well as the intermediate values of the first and second predicted frames, respectively $\hat{F}_{t-1}^i$

and $\hat{F}_{t+1}^i$ from iteration $i-1$. Specifically, the distortion D may be computed according to equation (6) as follows:

$$D = \sum_{(m,n) \in S} \left\| \hat{f}_{t-1}^{i+1}(m,n) - \hat{f}_{t-1}^i(m,n) \right\|^2 + \sum_{(m,n) \in S} \left\| \hat{f}_{t+1}^{i+1}(m,n) - \hat{f}_{t+1}^i(m,n) \right\|^2 + \sum_{(m,n) \in S} \left\| \hat{f}_t^{i+1}(m,n) - f_t(m,n) \right\|^2 \quad (6)$$

5 The distortion D is actually the minimum of the sum $\left\| \hat{F}_{t-1}^i - \hat{F}_{t-1}^{i+1} \right\|^2 + \left\| \hat{F}_{t+1}^i - \hat{F}_{t+1}^{i+1} \right\|^2 + \left\| F_t - \hat{F}_t^{i+1} \right\|^2$ using the optimal weight components.

The distortion takes into account the distortion between the intermediate values of the predicted frames from one iteration to the other, as well as the distortion between the prediction of the second frame and the second frame
10 itself.

If D is smaller than a preset threshold in act 560, or i is larger than the pre-defined maximum iteration number in act 520, the iteration is terminated and the current weight vector $\vec{W}^i$ is set to be the ultimate weight vector of the STAR model. Thus, the interpolated pixels values generated by $\vec{W}^i$ are
15 considered to be the final interpolated pixels values that defined the first and second predicted frames. Otherwise, i is increased by 1 and the self-feedback weight training algorithm is moved to the next iteration.

Figure 6 describes an illustrative embodiment of the self-feedback weight
20 training process according to the present invention. The interpolated pixels in frames $t-1$ and $t+1$ are used to approximate the actual pixels in the original low rate frame t . As shown in Figure 6, the actual pixels within temporal layer t may be approximated according to equation (4) described here above.

25 Figure 7 describes the MCI method. MCI is a very common FRUC method, which is performed block by block. It is composed of two processes: motion estimation and motion compensation. The motion estimation is performed by a bidirectional motion estimation method, which is shown as

Figure 7. Starting from the centre of the current to-be-interpolated block, bidirectional motion estimation finds two matching blocks in the previous and following frames in a reverse way. Next, the pixels in the to-be-interpolated frames are computed by motion compensation according to equation (7) as follows:

$$F_{t-1}^0(x, y) = \frac{1}{2} \left(F_{t-2}(x + v_x, y + v_y) + F_t(x - v_x, y - v_y) \right) \quad (7.1)$$

$$F_{t+1}^0(x, y) = \frac{1}{2} \left(F_t(x + v_x, y + v_y) + F_{t+2}(x - v_x, y - v_y) \right) \quad (7.2)$$

where (v_x, v_y) is the motion vector found in the bidirectional motion estimation process.

Figure 8 describes an example of simulation performed using the method according to the invention, i.e. the STAR model.

Various standard video sequences, with varying sizes (QCIF, CIF), have been tested in an attempt to shed some light on the quality of interpolated frames. To compare the proposed model with other FRUC algorithms, every other frame from test sequences listed in Figure 8 is skipped and interpolated by STAR model as well as other FRUC methods. In the self-feedback weight training method, the interpolated frames yielded by the MCI-FRUC method are set to be the initial pixels within frames $t-1$ and $t+1$ before the iteration starts, it a similar way to what is done in the present method in relation with Figure 7. The interpolated frames are then compared with the accurate ones within original sequences.

The peak signal-noise-ratio (PSNR) is used as an objective image quality measure to evaluate the performances of the interpolated frames yielded by MCI, OBMC-MCI, AOBMC-MCI, and STAR model. The PSNR gains of OBMC-MCI, AOBMC-MCI, as well as STAR model compared with the MCI method, averaged over 50 frames within all the test sequences, are depicted in **Figure 8**.

It may be seen that the gain of STAR model is very significant, especially for QCIF

sequence *mobile* and CIF sequence *Flower*, the gain can be up to 2.832dB and 1.824dB, respectively. This is because the *mobile* sequence is composed of camera motions such as zooming and panning and consequently traditional FRUC methods, e.g. MCI, OBMC-MCI and AOBMC-MCI, prove to be inefficient to represent the camera motions. *Flower* sequence is full of non-rigid objects and consequently could not be approximated well based on the assumption that all the pixels within one block have unique motion, which are the foundation of traditional FRUC methods. For the majority of the test sequences, OBMC and AOBMC achieve higher PSNR than those of MCI method due to the application of multi-hypothesis and thus can alleviate the artifacts when the shape of the objects are not aligned with the blocks. AOBMC achieves better performance than OBMC, since it is capable of adjusting the coefficients adaptively according to the reliability of neighbouring blocks. However, the improvement of AOBMC is rather poor for some sequences, e.g. *Miss American* and *Akiyo*. In these two sequences, the motions of neighbouring blocks are very similar to that of the current block, thus the superiority is impaired. Since the proposed STAR model fully exploits the correlations both in spatial and temporal domains within adjacent frames, it outperforms MCI, OBMC-MCI, AOBMC-MCI, for all the test sequences, no matter what the format is.

The method according to the invention may be carried out by a device. Said device may be comprised in a system. As described in Figure 9, said system may comprise for instance an emitting device 900 (e.g. a encoder), a receiving device 920 (e.g. a decoder) and a network 910 for communicating between the emitting device 900 and the receiving device 920. The method according to the invention may be implemented either on the emitting device 900 or on the receiving device 920.

Claims

1. A method for adding frames in a digital video flow of frames, the added frames comprising a first predicted frame added between a first and a second frames of the digital video flow, and a second predicted frame added between a second and a third frames of said digital video flow, the method further comprising the acts of:
 - a) initializing an intermediate estimate of the first and second predicted frames,
 - and for each intermediate estimate of the first and second predicted frames:
 - b1) defining a new intermediate estimate of the first predicted frame using a first interpolation function applied to the first and second frames,
 - b2) defining a new intermediate estimate of the second predicted frame using a second interpolation function applied to the second and third frames,
 - b3) defining a prediction of the second frame with a third interpolation function applied to the intermediate estimate of the first and second predicted frames,
 - b4) optimizing the interpolation functions while minimizing the distortion between the new intermediate estimate and the intermediate estimate of the predicted frames and between the second frame and its prediction,
 - b5) computing the intermediate estimates of the first and second predicted frames using the optimized first and second interpolation functions,
 - the first and second predicted frames being set to the intermediate estimates of the first and second predicted frames when the distortion is smaller than a given threshold or a given number of intermediate estimates of the first and second predicted frames has been computed.
2. The method according to claim 1, wherein the first, second and third interpolation function are identical.

3. The method according to one of the previous claims, wherein in acts b1), b2) and b3), the defining act is carried pixel by pixel, and wherein the interpolation function is also applied to the already defined pixels.
- 5 4. A interpolating device for adding frames in a digital video flow of frames, the added frames comprising a first predicted frame added between a first and a second frames of the digital video flow, and a second predicted frame added between a second and a third frames of the digital video flow, said device being operable to receive said digital video flow of frames, the device being
10 further operable to:
 - a) initialize an intermediate estimate of the first and second predicted frames, and for each intermediate estimate of the first and second predicted frames:
 - b1) define a new intermediate estimate of the first predicted frame using a first interpolation function applied to the first and second frames,
 - 15 b2) define a new intermediate estimate of the second predicted frame using a second interpolation function applied to the second and third frames,
 - b3) define a prediction of the second frame with a third interpolation function applied to the intermediate estimate of the first and second predicted frames,
 - b4) optimize the interpolation functions while minimizing the distortion
20 between the new intermediate estimate and the intermediate estimate of the predicted frames and between the second frame and its prediction,
 - b5) compute the intermediate estimates of the first and second predicted frames using the optimized first and second interpolation functions,
 - the first and second predicted frames being set to the intermediate estimates
25 of the first and second predicted frames when the distortion is smaller than a given threshold or a given number of intermediate estimates of the first and second predicted frames has been computed.
5. The interpolating device according to claim 4, wherein the first, second and
30 third interpolation function are identical.

6. The interpolating device according to any of the previous claims 4 and 5, wherein said device is further operable to carry out acts b1), b2) and b3) pixel by pixel, and wherein the interpolation function is also applied to the already defined pixels.
7. A system for adding frames in a digital video flow of frames, the added frames comprising a first predicted frame added between a first and a second frames of the digital video flow, and a second predicted frame added between a second and a third frames of said digital video flow, said system comprising:
- an emitting device operable to send said digital video flow,
 - a receiving device operable to:
 - a) initialize an intermediate estimate of the first and second predicted frames,
 - and for each intermediate estimate of the first and second predicted frames:
 - b1) define a new intermediate estimate of the first predicted frame using a first interpolation function applied to the first and second frames,
 - b2) define a new intermediate estimate of the second predicted frame using a second interpolation function applied to the second and third frames,
 - b3) define a prediction of the second frame with a third interpolation function applied to the intermediate estimate of the first and second predicted frames,
 - b4) optimize the interpolation functions while minimizing the distortion between the new intermediate estimate and the intermediate estimate of the predicted frames and between the second frame and its prediction,

b5) compute the intermediate estimates of the first and second predicted frames using the optimized first and second interpolation functions,

the first and second predicted frames being set to the intermediate estimates of the first and second predicted frames when the distortion is smaller than a given threshold or a given number of intermediate estimates of the first and second predicted frames has been computed;

- a network operable to communicate between the emitting device and the receiving device.

10

8. The system according to claim 7, wherein the first, second and third interpolation function are identical.

15

9. The system according to any of the previous claims 7 and 8, wherein said system is further operable to carry the defining acts b1), b2) and b3) pixel by pixel, and wherein the interpolation function is also applied to the already defined pixels.

20

10. A computer program providing computer executable instructions stored on a computer readable medium, which when loaded on to a data processor causes the data processor to perform a method for adding frames in a digital video flow of frames according to claims 1 to 3.

25

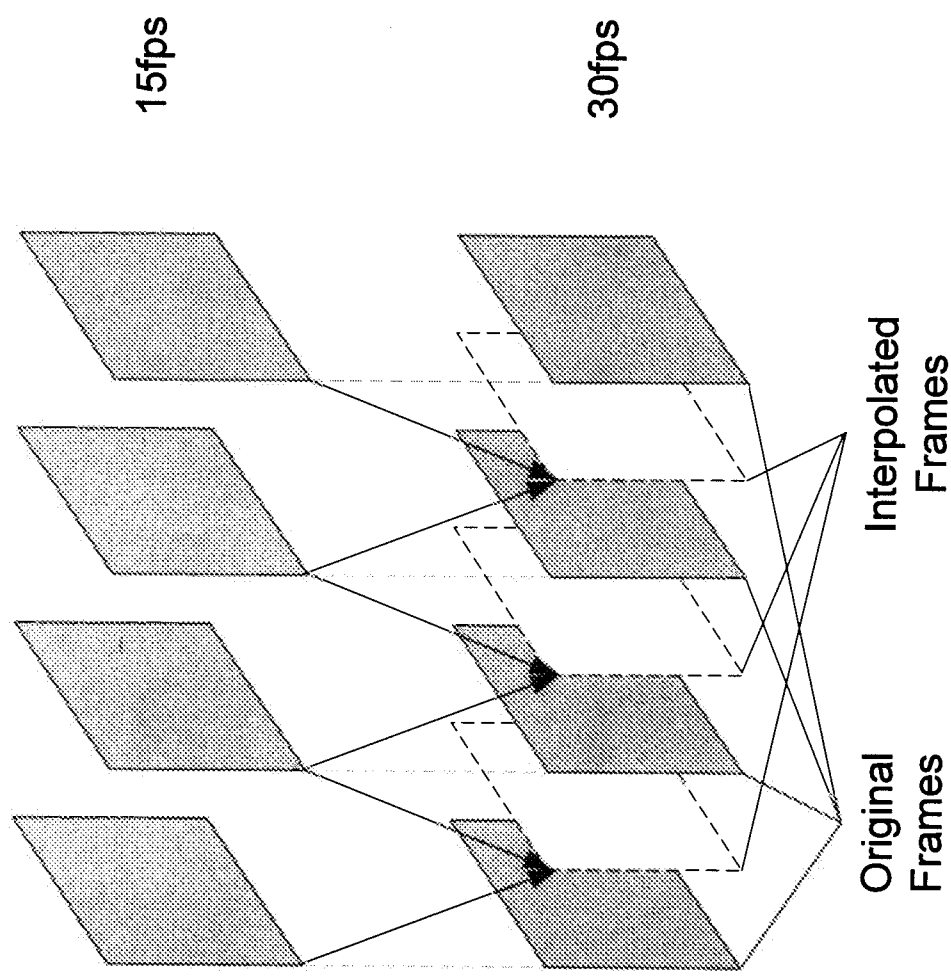


Figure 1

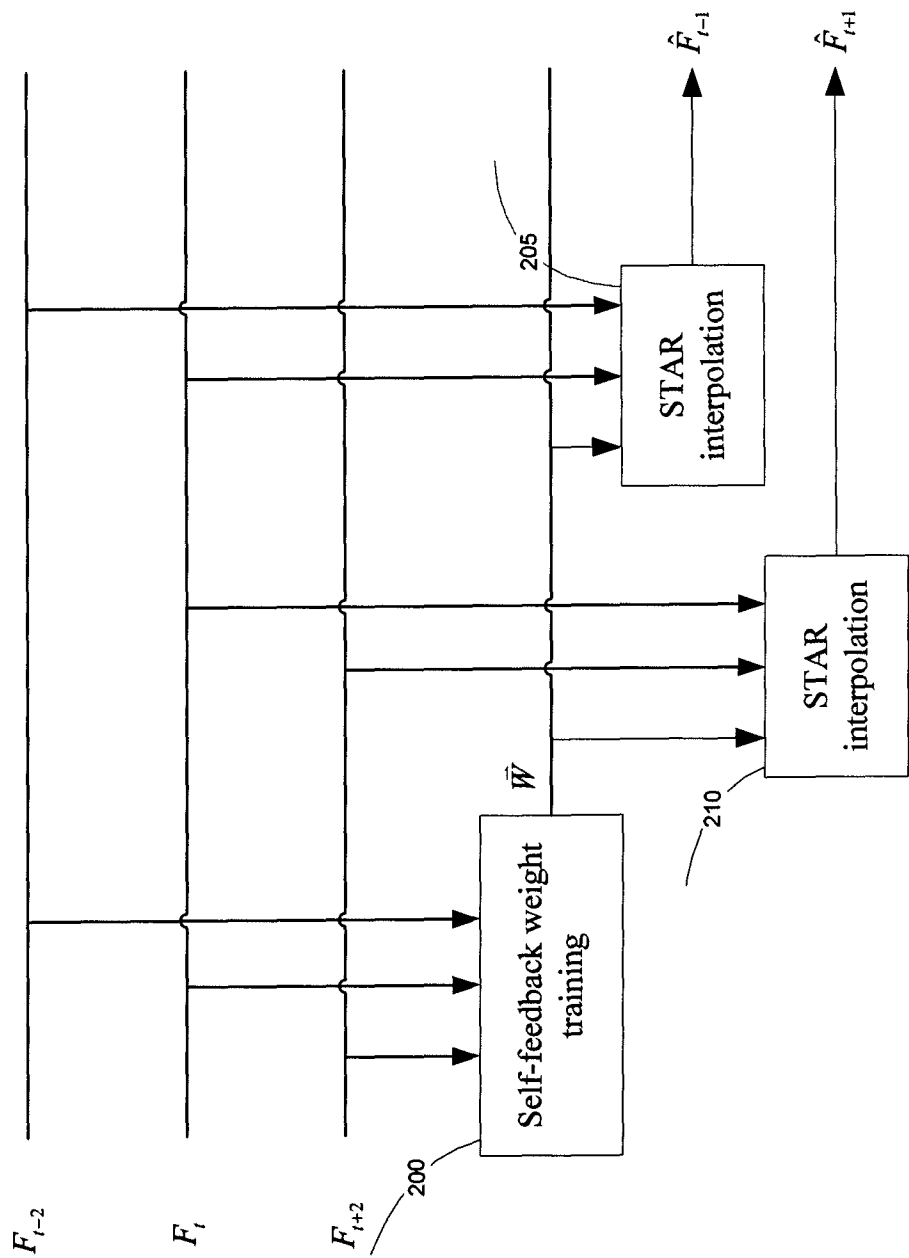


Figure 2

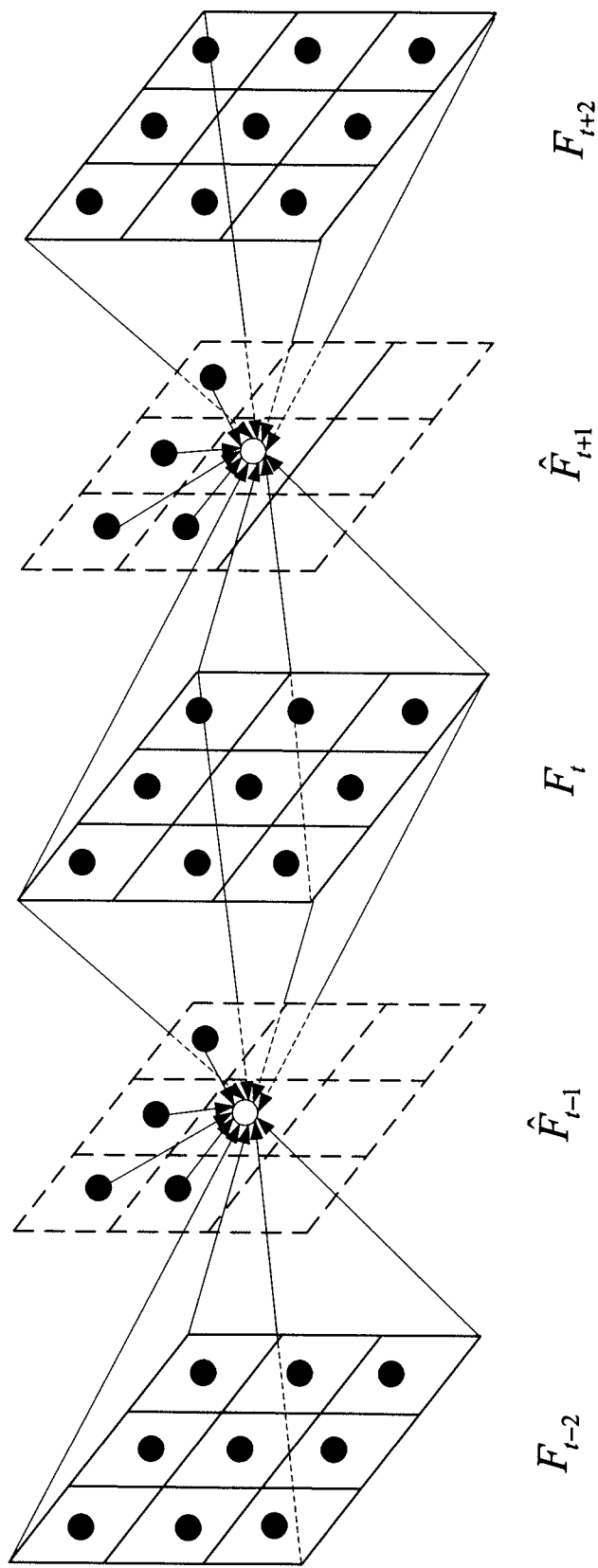
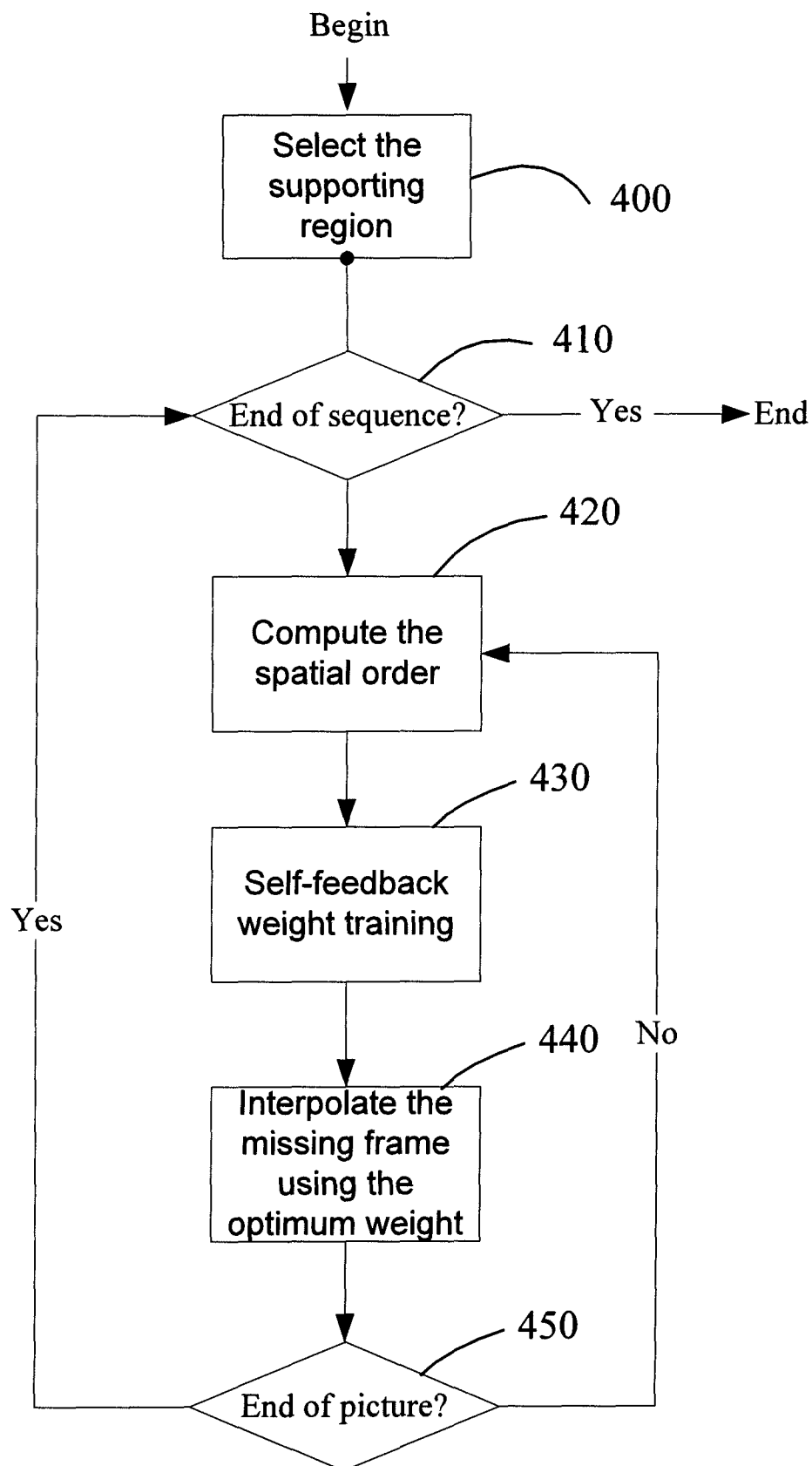


Figure 3

**Figure 4**

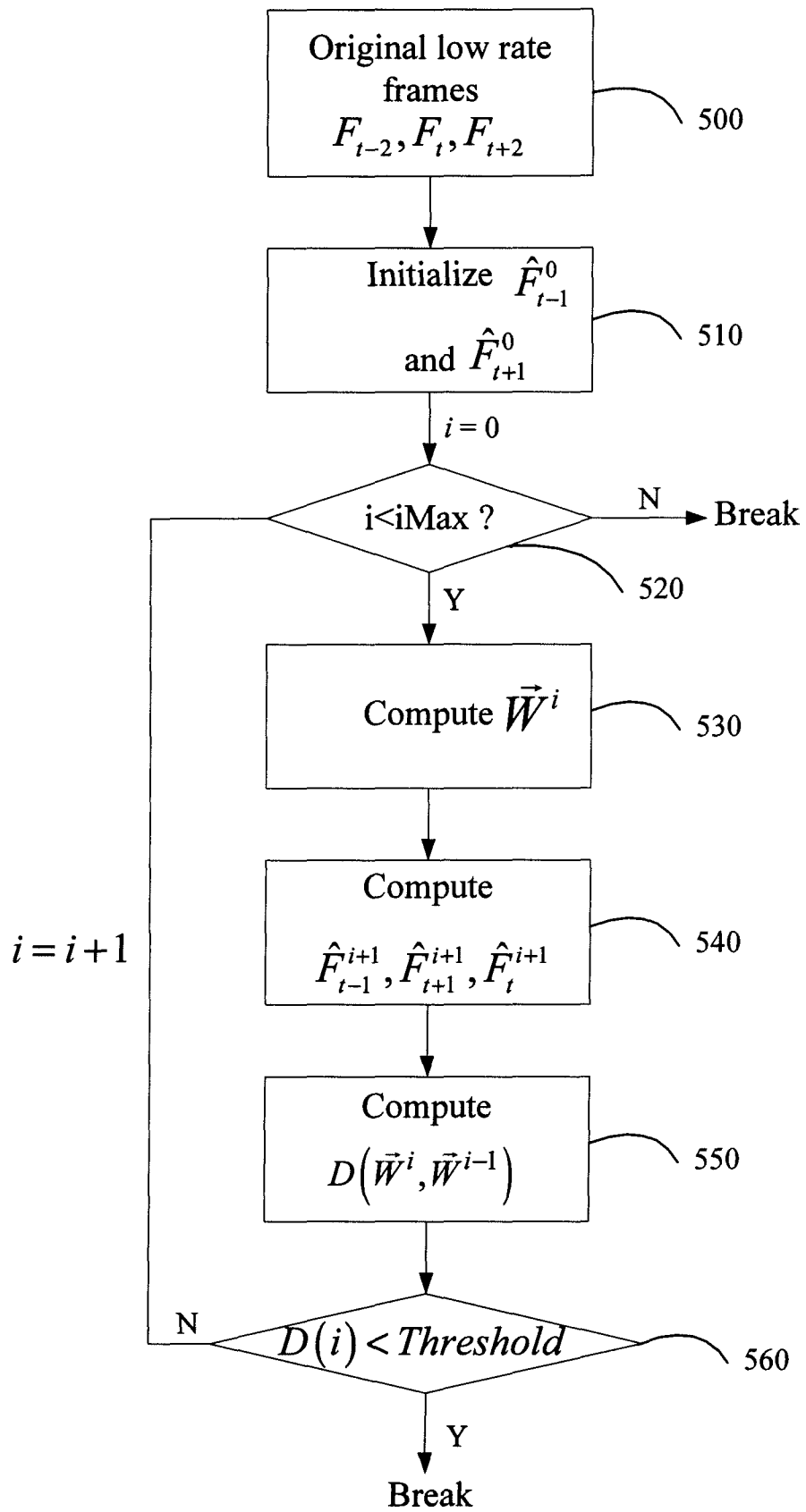


Figure 5

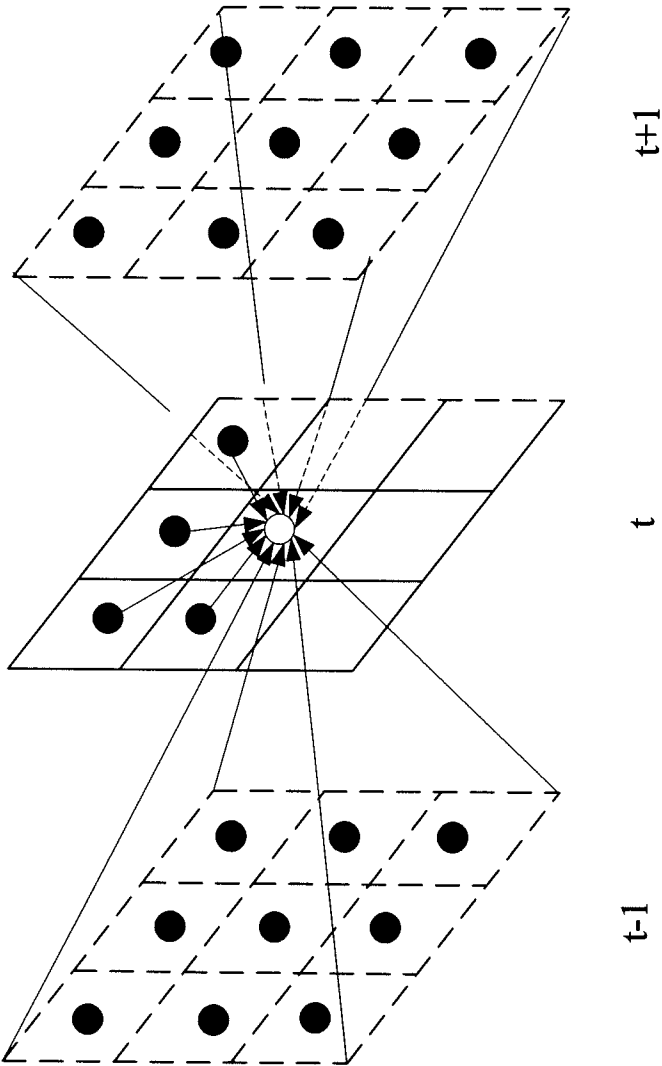


Figure 6

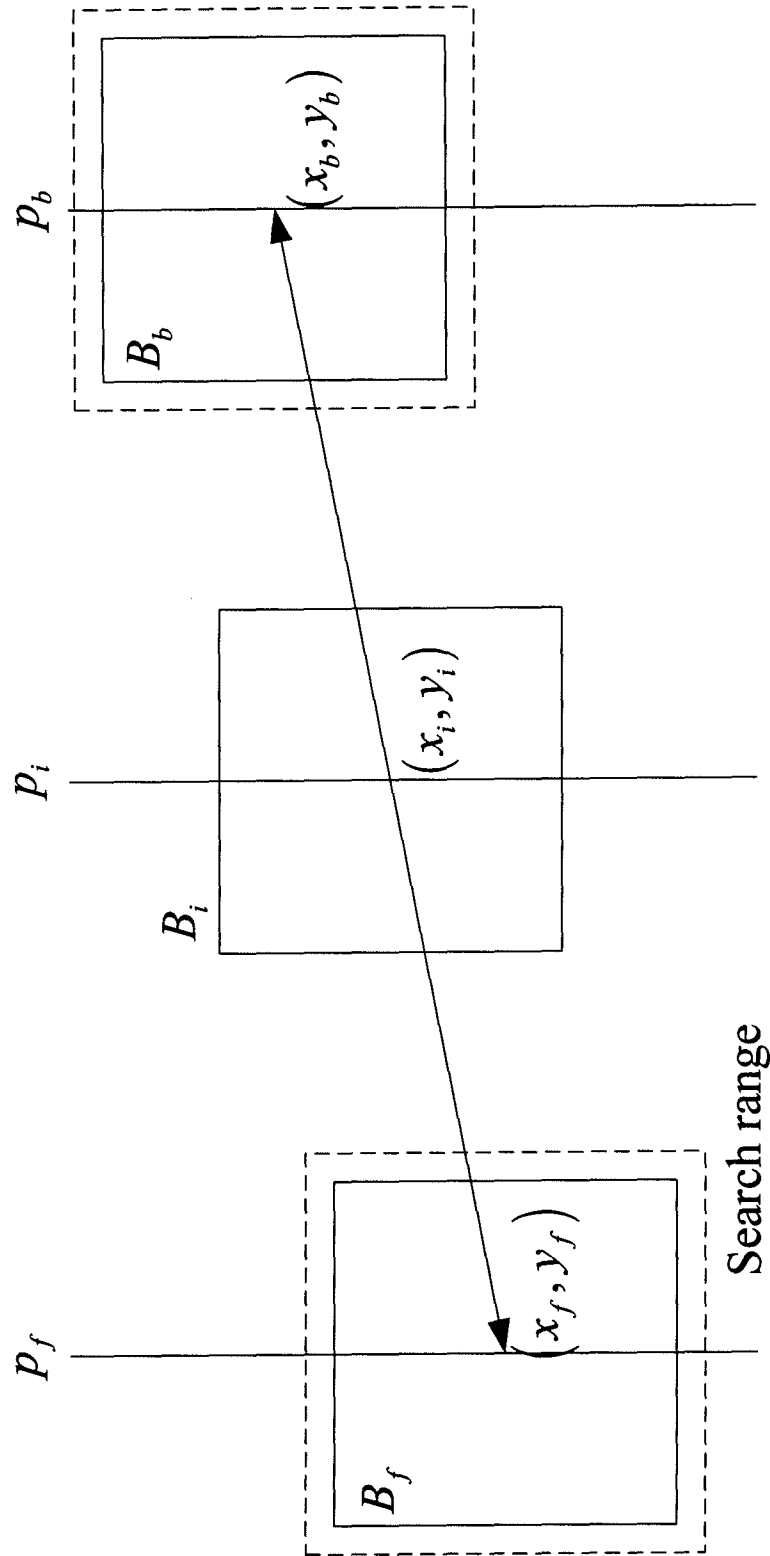


Figure 7

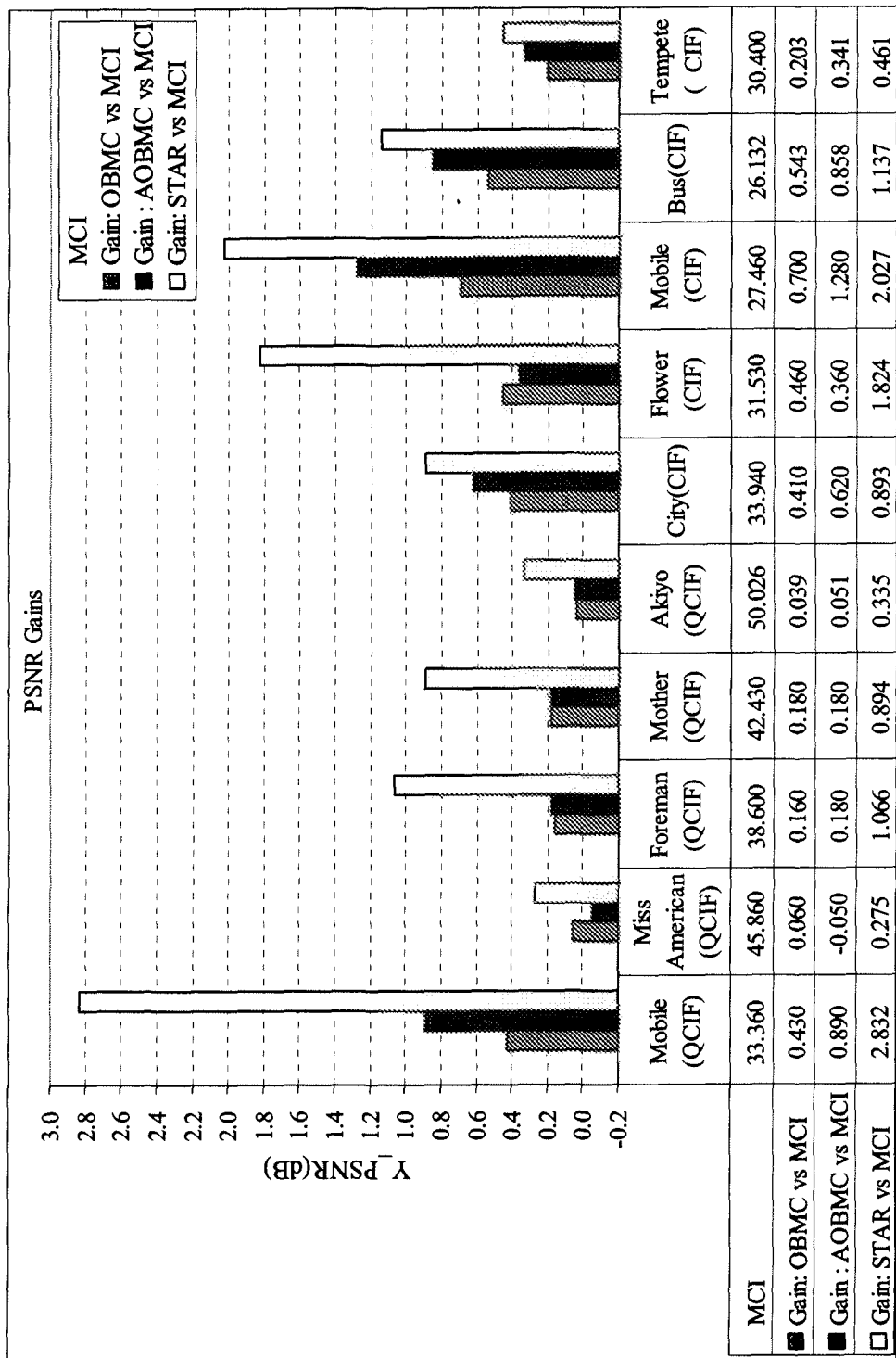


Figure 8

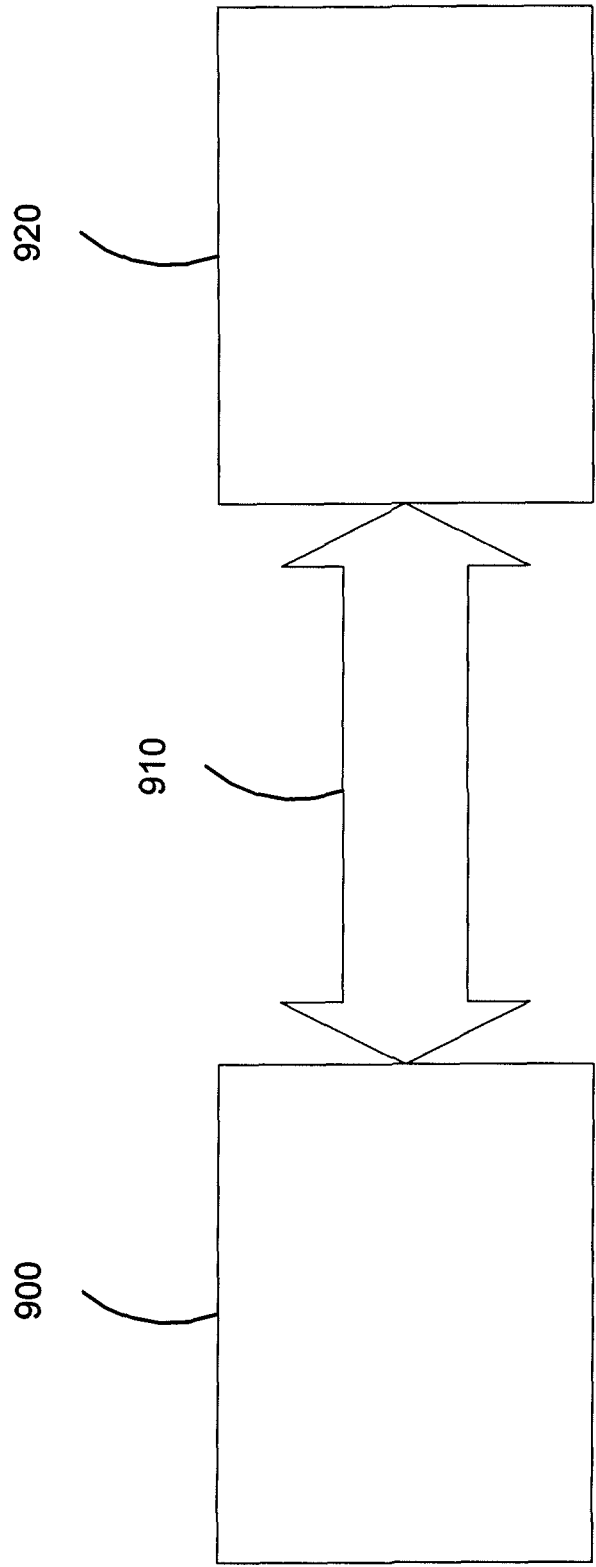


Figure 9

INTERNATIONAL SEARCH REPORT

International application No
PCT/IB2009/055097

A. CLASSIFICATION OF SUBJECT MATTER
INV. H04N7/01

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
H04N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	YONGBING ZHANG ET AL: "A Spatio-Temporal Autoregressive Frame Rate Up Conversion Scheme" IMAGE PROCESSING, 2007. ICIP 2007. IEEE INTERNATIONAL CONFERENCE ON, IEEE, PI, 1 September 2007 (2007-09-01), pages I-441, XP031157773 ISBN: 978-1-4244-1436-9 2.1 Parameter estimation; pages I-442; figure 2 -----	1-10

☐ Further documents are listed in the continuation of Box C.

☐ See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"8" document member of the same patent family

Date of the actual completion of the international search

10 February 2010

Date of mailing of the international search report

17/02/2010

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Beaudoin, Olivier