



- (51) International Patent Classification:
H04N 7/26 (2006.01) H04N 7/50 (2006.01)
H04N 7/36 (2006.01)
- (21) International Application Number:
PCT/US2009/002177
- (22) International Filing Date:
7 April 2009 (07.04.2009)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
61/044,193 11 April 2008 (11.04.2008) US
- (71) Applicant (for all designated States except US): THOMSON LICENSING [FR/FR]; 46, Quai A. Le Gallo, F-92100 Boulogne-Billancourt (FR).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): ZHENG, Yunfei [CN/US]; 5402 Ravens Crest Dr., Plainsboro, NJ 08536 (US). DIVORRA ESCODA, Oscar [ES/ES]; 171-177 Av. Madrid, E-08028 Barcelona (ES). YIN, Peng [CN/US]; 49 Thoreau Drive, Plainsboro, NJ 08536 (US). SOLE, Joel [ES/US]; 1006 Quail Ridge Drive, Plainsboro, NJ 08536 (US).
- (74) Agents: SHEDD, Robert, D. et al.; Thomson Licensing LLC, Two Independence Way, Suite 200, Princeton, NJ 08540 (US).

- (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, 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, 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, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

- with international search report (Art. 21(3))
- 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))

(54) Title: METHODS AND APPARATUS FOR TEMPLATE MATCHING PREDICTION (TMP) IN VIDEO ENCODING AND DECODING

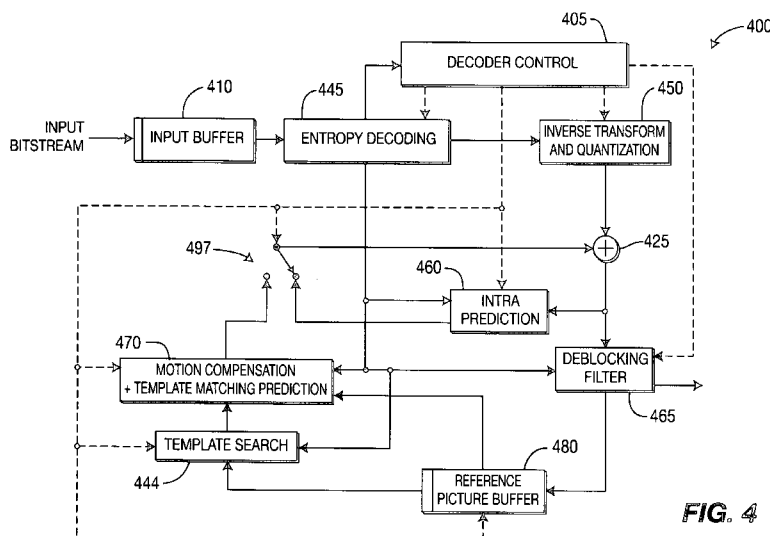


FIG. 4

(57) Abstract: Methods and apparatus are provided for template matching prediction in video encoding and decoding. An apparatus includes an encoder (300) for encoding a target block in a picture by determining at least one predictor for the target block and respectively utilizing the at least one predictor as a template to search for at least one additional predictor for the target block.

WO 2009/126260 A1

METHODS AND APPARATUS FOR TEMPLATE MATCHING PREDICTION (TMP) IN VIDEO ENCODING AND DECODING

CROSS-REFERENCE TO RELATED APPLICATIONS

5 This application claims the benefit of U.S. Provisional Application Serial No. 61/044,193 (Attorney Docket No. PU080048), filed 11 April 2008, which is incorporated by reference herein in its entirety.

TECHNICAL FIELD

10 The present principles relate generally to video encoding and decoding and, more particularly, to methods and apparatus for template matching prediction (TMP) in video encoding and decoding.

BACKGROUND

15 Template matching prediction (TMP), used in video coding and decoding schemes, can create gains in coding efficiency for inter-prediction and/or intra-prediction by avoiding transmission of motion vectors/displaced vectors. However, the prediction performance of template matching prediction is highly dependent on the correlation between the target block and its corresponding template. It is well
20 known in the art to use the reconstructed pixels surrounding the target block, on the top and/or to the left, as shown in FIG. 1. Turning to FIG. 1, an example of an implicit motion estimation scheme involving template matching prediction is indicated generally by the reference numeral 100. The implicit motion estimation scheme 100 involves a reconstructed reference frame 110 having a search region 111, a
25 prediction 112 within the search region 111, and a neighborhood 113 with respect to the prediction 112. The implicit motion estimation scheme 100 also involves a current frame 150 having a target block 151, a template 152 with respect to the target block 151, and a reconstructed region 153. However, utilizing template matching prediction in such a manner lessens encoder and decoder performance,
30 especially when the template has a lower correlation with the targeted signal.

Inter-Prediction

In video coding, inter-prediction is extensively employed to reduce temporal redundancy between the target frame and reference frames. Motion estimation and

compensation are components for inter-prediction. In general, we can classify motion models and corresponding motion estimation techniques into two categories. The first category is the so-called explicit motion estimation, or explicit motion model, where motion vectors are explicitly transmitted. The second category is the so-called implicit motion estimation, where motion information is estimated based on reconstructed pixels, so that no motion vectors are required to be transmitted.

The first category, based on block matching techniques, as shown in FIG. 2, is well adopted in current video coding standards such as, for example, the International Organization for Standardization/International Electrotechnical Commission (ISO/IEC) Moving Picture Experts Group-4 (MPEG-4) Part 10 Advanced Video Coding (AVC) standard/International Telecommunication Union, Telecommunication Sector (ITU-T) H.264 Recommendation (hereinafter the "MPEG-4 AVC Standard"). Turning to FIG. 2, an example of an explicit motion estimation scheme involving block matching is indicated generally by the reference numeral 200. The explicit motion estimation scheme 200 involves a reconstructed reference frame 210 having a search region 201 and a prediction 202 within the search region 201. The implicit motion estimation scheme 200 also involves a current frame 250 having a target block 251 and a reconstructed region 252. A motion vector MV is used to denote the motion between the target block 251 and the prediction 202. The first category is usually performed in two steps. The motion vectors between the target (current) block and the reference frames are estimated. Then the motion information is coded and explicitly sent to the decoder. At the decoder, the motion information is decoded and used to predict the target block from previously decoded reconstructed reference frames.

The second category refers to the class of motion estimation methods that do not code motion information explicitly in the bitstream. Instead, the same motion information derivation is performed at the decoder as that performed at the encoder. One practical implicit motion estimation scheme was to use a spatial-temporal autoregressive model, where the least-square prediction (LSP) is applied. Another common and popular approach is to use a patch-based model, such as the template matching prediction approach shown in FIG. 1.

In general, explicit motion estimation can be more accurate, especially for fast and unnatural motions, but at the cost of coding bits for motion vectors. Implicit motion estimation is more suitable for slow and natural motion, but it places a

complexity burden at the decoder since it has to compute the motion information as the encoder does.

MPEG-4 AVC Standard Inter-Prediction

5 The MPEG-4 AVC Standard uses tree-structured hierarchical macroblock partitions. Inter-coded 16x16 pixel macroblocks may be broken into macroblock partitions of sizes 16x8, 8x16, or 8x8. Macroblock partitions of 8x8 pixels are also known as sub-macroblocks. Sub-macroblocks may also be broken into sub-macroblock partitions of sizes 8x4, 4x8, and 4x4. An encoder may select how to
10 divide the macroblock into partitions and sub-macroblock partitions based on the characteristics of a particular macroblock, in order to maximize compression efficiency and subjective quality.

 Multiple reference pictures may be used for inter-prediction, with a reference picture index coded to indicate which of the multiple reference pictures is used. For
15 P pictures (or P slices), only single directional prediction is used, and the allowable reference pictures are managed in List 0. In B pictures (or B slices), two lists of reference pictures are managed, List 0 and List 1. In B pictures (or B slices), single directional prediction using either List 0 or List 1 is allowed, or bi-prediction using both List 0 and List 1 is allowed. When bi-prediction is used, the List 0 and the List 1
20 predictors are averaged together to form a final predictor.

 Each macroblock partition may have independent reference picture indices, prediction type (List 0, List 1, bipred), and an independent motion vector. Each sub-macroblock partition may have independent motion vectors, but all sub-macroblock partitions in the same sub-macroblock use the same reference picture index and
25 prediction type.

 In the MPEG-4 AVC Standard Joint Model (JM) reference software, a Rate-Distortion Optimization (RDO) framework is used for mode decision. For inter modes, motion estimation is separately considered from mode decision. Motion estimation is first performed for all block types of inter modes, then the mode
30 decision is made by comparing the cost of each inter mode and intra mode. The mode with the minimal cost is selected as the best mode. The following modes relate to a P-frame and a B-frame, respectively.

$$\begin{aligned}
 \text{P-frame: } & \text{MODE} \in \left\{ \text{INTRA } 4 \times 4, \text{ INTRA } 16 \times 16, \text{ SKIP}, \right. \\
 & \left. 16 \times 16, 16 \times 8, 8 \times 16, 8 \times 8, 8 \times 4, 4 \times 8, 4 \times 4 \right\}, \\
 \text{B-frame: } & \text{MODE} \in \left\{ \text{INTRA } 4 \times 4, \text{ INTRA } 16 \times 16, \text{ DIRECT}, \right. \\
 & \left. 16 \times 16L0, 16 \times 8L0, 8 \times 16L0, 8 \times 8L0, 8 \times 4L0, \right. \\
 & \left. 4 \times 8L0, 4 \times 4L0, 16 \times 16L1, 16 \times 8L1, 8 \times 16L1, \right. \\
 & \left. 8 \times 8L1, 8 \times 4L1, 4 \times 8L1, 4 \times 4L1, 16 \times 16BI, 16 \times 8BI, 8 \times 16BI, \right. \\
 & \left. 8 \times 8BI, 8 \times 4BI, 4 \times 8BI, 4 \times 4BI \right\},
 \end{aligned}$$

Template Matching

5 Template matching prediction is one way to predict target pixels without sending motion vectors. Given a target block of a frame, a target pixel in the block is determined by finding an optimum pixel from a set of reference samples, where the adjacent pixels of the optimum pixels have the highest correlation with those of the target pixels. Those adjacent pixels of the target pixels are called a template. The

10 prior art shows that the template is always taken from reconstructed surrounding pixels of the target pixels, as shown in FIG. 1. The template matching prediction can be applied for both intra and inter prediction.

In the case of inter-prediction, the template matching process can be seen as a motion vector search at the decoder side. Here, template matching is performed

15 very similar to traditional motion estimation techniques. Motion vectors are evaluated by calculating a cost function for accordingly displaced template-shaped regions in the reference frames. The best motion vector for the template is then used to predict the target area. Only those areas of the image where already a reconstruction or at least a prediction signal exists are accessed for the search.

20 Thus, the decoder is able to execute the template matching process and predict the target area without additional side information.

Template matching can predict pixels in a target block without transmission of motion vectors. It is expected that the prediction performance of template matching prediction is comparable to that of the block matching scheme if the correlation

25 between a target block and its template is high. In the prior art, the template is always taken from the reconstructed spatial neighboring pixels of the target pixels. The neighboring pixels sometimes have low correlations with the target pixels, thus the performance of template matching prediction can be lower than traditional block matching scheme.

Multiple-hypothesis Motion Compensated Prediction (MCP)

Multiple-hypothesis is the combination of multiple prediction candidates to make the prediction signals of a target block less noisy, or more correlated to the target block. Bi-prediction is the most common case of multiple-hypothesis. In the prior art, multiple-hypothesis motion compensated prediction can be classified into two categories, along with explicit prediction and implicit prediction. In the first category, all motion vectors are explicitly transmitted for each hypothesis motion compensated prediction. In the second category, no motion vector is embedded in the bitstream, such as template matching averaging. The disadvantage of the first approach is the bitrate overhead for motion. The disadvantage of the second approach is the inaccuracy introduced by the outliers that are possibly involved because the matching process is based on the neighboring pixels.

SUMMARY

These and other drawbacks and disadvantages of the prior art are addressed by the present principles, which are directed to methods and apparatus for template matching prediction (TMP) in video encoding and decoding.

According to an aspect of the present principles, there is provided an apparatus. The apparatus includes an encoder for encoding a target block in a picture by determining at least one predictor for the target block and respectively utilizing the at least one predictor as a template to search for at least one additional predictor for the target block.

According to another aspect of the present principles, there is provided a method. The method includes encoding a target block in a picture by determining at least one predictor for the target block and respectively utilizing the at least one predictor as a template to search for at least one additional predictor for the target block.

According to yet another aspect of the present principles, there is provided an apparatus. The apparatus includes a decoder for decoding a target block in a picture by determining at least one predictor for the target block and respectively utilizing the at least one predictor as a template to search for at least one additional predictor for the target block.

According to a still further aspect of the present principles, there is provided a method. The method includes decoding a target block in a picture by determining at least one predictor for the target block and respectively utilizing the at least one predictor as a template to search for at least one additional predictor for the target block.

These and other aspects, features and advantages of the present principles will become apparent from the following detailed description of exemplary embodiments, which is to be read in connection with the accompanying drawings.

10 BRIEF DESCRIPTION OF THE DRAWINGS

The present principles may be better understood in accordance with the following exemplary figures, in which:

FIG. 1 is a diagram showing an example of an implicit motion estimation scheme involving template matching prediction;

15 FIG. 2 is a diagram showing an example of an explicit motion estimation scheme involving block matching;

FIG. 3 is a block diagram showing an exemplary video encoder to which the present principles may be applied, in accordance with an embodiment of the present principles;

20 FIG. 4 is a block diagram showing an exemplary video decoder to which the present principles may be applied, in accordance with an embodiment of the present principles;

FIG. 5 is a diagram showing an example of template matching prediction (TMP) scheme combined with single prediction (uni-prediction), in accordance with an embodiment of the present principles;

FIG. 6 is a flow diagram showing a video encoding method using template matching prediction with single reference list prediction (uni-prediction), in accordance with an embodiment of the present principles;

30 FIG. 7 is a flow diagram showing a video decoding method using template matching prediction with single reference list prediction (uni-prediction), in accordance with an embodiment of the present principles;

FIG. 8 is a diagram showing an example of template matching prediction combined with bi-prediction, in accordance with an embodiment of the present principles;

FIG. 9 is a flow diagram showing a video encoding method using template matching prediction with bi-prediction, in accordance with an embodiment of the present principles;

5 FIG. 10 is a flow diagram showing a video decoding method using template matching prediction with bi-prediction, in accordance with an embodiment of the present principles;

FIG. 11 is a flow diagram showing another video encoding method using template matching prediction with bi-prediction, in accordance with an embodiment of the present principles;

10 FIG. 12 is a flow diagram showing another video decoding method using template matching prediction with bi-prediction, in accordance with an embodiment of the present principles;

FIG. 13 is a diagram showing an example of template matching prediction with multiple reference pictures, in accordance with an embodiment of the present principles, in accordance with an embodiment of the present principles;

15 FIG. 14 is a flow diagram showing a video encoding method using template matching prediction with multiple reference pictures, in accordance with an embodiment of the present principles;

20 FIG. 15 is a flow diagram showing a video decoding method using template matching prediction with multiple reference pictures, in accordance with an embodiment of the present principles;

FIG. 16 is a flow diagram showing another video encoding method using template matching prediction for intra prediction, in accordance with an embodiment of the present principles; and

25 FIG. 17 is a flow diagram showing another video decoding method using template matching prediction with multiple reference pictures, in accordance with an embodiment of the present principles.

DETAILED DESCRIPTION

30 The present principles are directed to methods and apparatus for template matching prediction (TMP) in video encoding and decoding.

The present description illustrates the present principles. It will thus be appreciated that those skilled in the art will be able to devise various arrangements

that, although not explicitly described or shown herein, embody the present principles and are included within its spirit and scope.

All examples and conditional language recited herein are intended for pedagogical purposes to aid the reader in understanding the present principles and the concepts contributed by the inventor(s) to furthering the art, and are to be construed as being without limitation to such specifically recited examples and conditions.

Moreover, all statements herein reciting principles, aspects, and embodiments of the present principles, as well as specific examples thereof, are intended to encompass both structural and functional equivalents thereof. Additionally, it is intended that such equivalents include both currently known equivalents as well as equivalents developed in the future, i.e., any elements developed that perform the same function, regardless of structure.

Thus, for example, it will be appreciated by those skilled in the art that the block diagrams presented herein represent conceptual views of illustrative circuitry embodying the present principles. Similarly, it will be appreciated that any flow charts, flow diagrams, state transition diagrams, pseudocode, and the like represent various processes which may be substantially represented in computer readable media and so executed by a computer or processor, whether or not such computer or processor is explicitly shown.

The functions of the various elements shown in the figures may be provided through the use of dedicated hardware as well as hardware capable of executing software in association with appropriate software. When provided by a processor, the functions may be provided by a single dedicated processor, by a single shared processor, or by a plurality of individual processors, some of which may be shared. Moreover, explicit use of the term "processor" or "controller" should not be construed to refer exclusively to hardware capable of executing software, and may implicitly include, without limitation, digital signal processor ("DSP") hardware, read-only memory ("ROM") for storing software, random access memory ("RAM"), and non-volatile storage.

Other hardware, conventional and/or custom, may also be included. Similarly, any switches shown in the figures are conceptual only. Their function may be carried out through the operation of program logic, through dedicated logic, through the interaction of program control and dedicated logic, or even manually, the

particular technique being selectable by the implementer as more specifically understood from the context.

In the claims hereof, any element expressed as a means for performing a specified function is intended to encompass any way of performing that function including, for example, a) a combination of circuit elements that performs that function or b) software in any form, including, therefore, firmware, microcode or the like, combined with appropriate circuitry for executing that software to perform the function. The present principles as defined by such claims reside in the fact that the functionalities provided by the various recited means are combined and brought together in the manner which the claims call for. It is thus regarded that any means that can provide those functionalities are equivalent to those shown herein.

Reference in the specification to "one embodiment" or "an embodiment" of the present principles, as well as other variations thereof, means that a particular feature, structure, characteristic, and so forth described in connection with the embodiment is included in at least one embodiment of the present principles. Thus, the appearances of the phrase "in one embodiment" or "in an embodiment", as well as any other variations, appearing in various places throughout the specification are not necessarily all referring to the same embodiment.

It is to be appreciated that the use of any of the following "/", "and/or", and "at least one of", for example, in the cases of "A/B", "A and/or B" and "at least one of A and B", is intended to encompass the selection of the first listed option (A) only, or the selection of the second listed option (B) only, or the selection of both options (A and B). As a further example, in the cases of "A, B, and/or C" and "at least one of A, B, and C", such phrasing is intended to encompass the selection of the first listed option (A) only, or the selection of the second listed option (B) only, or the selection of the third listed option (C) only, or the selection of the first and the second listed options (A and B) only, or the selection of the first and third listed options (A and C) only, or the selection of the second and third listed options (B and C) only, or the selection of all three options (A and B and C). This may be extended, as readily apparent by one of ordinary skill in this and related arts, for as many items listed.

Moreover, it is to be appreciated that while one or more embodiments of the present principles are described herein with respect to the MPEG-4 AVC standard, the present principles are not limited to solely this standard and, thus, may be utilized with respect to other video coding standards, recommendations, and

extensions thereof, including extensions of the MPEG-4 AVC standard, while maintaining the spirit of the present principles.

Further, it is to be appreciated that the terms "prediction", "predictor", and "hypothesis" are interchangeably used herein.

5 Also, it is to be appreciated that the term "block" as used herein refers to any of a macroblock, a macroblock partition, a sub-macroblock, and a sub-macroblock partition. Moreover, the term "block" may also refer to, for example, a wedge segmented from any of a macroblock, a macroblock partition, a sub-macroblock, and a sub-macroblock partition.

10 Additionally, it is to be appreciated that the phrases "explicit coding" and "explicit decoding" respectively refer to the cases where encoding or decoding is performed such that corresponding syntax relating to the encoding or decoding is included in a bitstream (e.g., the bitstream including the video content to which the syntax corresponds or a separate bitstream(s)). In contrast, the phrases "implicit coding" and "implicit decoding" respectively refer to the cases where encoding or decoding is performed such that corresponding syntax relating to the encoding or decoding is not included in a bitstream (e.g., the bitstream including the video content to which the syntax corresponds or a separate bitstream(s)). Rather, the corresponding syntax may be derived from other syntax, and so forth.

20 Turning to FIG. 3, an exemplary video encoder to which the present principles may be applied is indicated generally by the reference numeral 300.

The video encoder 300 includes a frame ordering buffer 310 having an output in signal communication with a non-inverting input of a combiner 385. An output of the combiner 385 is connected in signal communication with a first input of a transformer and quantizer 325. An output of the transformer and quantizer 325 is connected in signal communication with a first input of an entropy coder 345 and a first input of an inverse transformer and inverse quantizer 350. An output of the entropy coder 345 is connected in signal communication with a first non-inverting input of a combiner 390. An output of the combiner 390 is connected in signal communication with a first input of an output buffer 335.

30 A first output of an encoder controller 305 is connected in signal communication with a second input of the frame ordering buffer 310, a second input of the inverse transformer and inverse quantizer 350, an input of a picture-type decision module 315, a first input of a macroblock-type (MB-type) decision module

320, a second input of an intra prediction module 360, a second input of a deblocking filter 365, a first input of a motion compensator and template matching predictor 370, a first input of a motion estimator 375, and a second input of a reference picture buffer 380.

5 A second output of the encoder controller 305 is connected in signal communication with a first input of a Supplemental Enhancement Information (SEI) inserter 330, a second input of the transformer and quantizer 325, a second input of the entropy coder 345, a second input of the output buffer 335, and an input of the Sequence Parameter Set (SPS) and Picture Parameter Set (PPS) inserter 340.

10 An output of the SEI inserter 330 is connected in signal communication with a second non-inverting input of the combiner 390.

A first output of the picture-type decision module 315 is connected in signal communication with a third input of a frame ordering buffer 310. A second output of the picture-type decision module 315 is connected in signal communication with a second input of a macroblock-type decision module 320.

15 An output of the Sequence Parameter Set (SPS) and Picture Parameter Set (PPS) inserter 340 is connected in signal communication with a third non-inverting input of the combiner 390.

An output of the inverse quantizer and inverse transformer 350 is connected in signal communication with a first non-inverting input of a combiner 319. An output of the combiner 319 is connected in signal communication with a first input of the intra prediction module 360 and a first input of the deblocking filter 365. An output of the deblocking filter 365 is connected in signal communication with a first input of a reference picture buffer 380. An output of the reference picture buffer 380 is connected in signal communication with a second input of the motion estimator 375, a third input of the motion compensator and template matching predictor 370, and a first input of a template searcher 344. An output of the template searcher 344 is connected in signal communication with a fourth input of the motion compensator and template matching predictor 370. A first output of the motion estimator 375 is connected in signal communication with a second input of the motion compensator and template matching predictor 370. A second output of the motion estimator 375 is connected in signal communication with a third input of the entropy coder 345. A third output of the motion estimator 375 is connected in signal communication with a second input of the template searcher 344. A third output of the encoder controller

305 is connected in signal communication with a third input of the template searcher 344.

An output of the motion compensator and template matching predictor 370 is connected in signal communication with a first input of a switch 397. An output of the intra prediction module 360 is connected in signal communication with a second input of the switch 397. An output of the macroblock-type decision module 320 is connected in signal communication with a third input of the switch 397. The third input of the switch 397 determines whether or not the "data" input of the switch (as compared to the control input, i.e., the third input) is to be provided by the motion compensator and template matching predictor 370 or the intra prediction module 360. The output of the switch 397 is connected in signal communication with a second non-inverting input of the combiner 319 and with an inverting input of the combiner 385.

A first input of the frame ordering buffer 310 and an input of the encoder controller 305 are available as input of the encoder 100, for receiving an input picture. Moreover, a second input of the Supplemental Enhancement Information (SEI) inserter 330 is available as an input of the encoder 300, for receiving metadata. An output of the output buffer 335 is available as an output of the encoder 300, for outputting a bitstream.

Turning to FIG. 4, an exemplary video decoder to which the present principles may be applied is indicated generally by the reference numeral 400.

The video decoder 400 includes an input buffer 410 having an output connected in signal communication with a first input of the entropy decoder 445. A first output of the entropy decoder 445 is connected in signal communication with a first input of an inverse transformer and inverse quantizer 450. An output of the inverse transformer and inverse quantizer 450 is connected in signal communication with a second non-inverting input of a combiner 425. An output of the combiner 425 is connected in signal communication with a second input of a deblocking filter 465 and a first input of an intra prediction module 460. A second output of the deblocking filter 465 is connected in signal communication with a first input of a reference picture buffer 480. An output of the reference picture buffer 480 is connected in signal communication with a first input of a template searcher 444. An output of the template searcher 444 is connected in signal communication with a second input of a motion compensator and template matching predictor 470.

A second output of the entropy decoder 445 is connected in signal communication with a third input of the motion compensator and template matching predictor 470, a first input of the deblocking filter 465, and a second input of the template searcher 444. A third output of the entropy decoder 445 is connected in signal communication with an input of a decoder controller 405. A first output of the decoder controller 405 is connected in signal communication with a second input of the entropy decoder 445. A second output of the decoder controller 405 is connected in signal communication with a second input of the inverse transformer and inverse quantizer 450. A third output of the decoder controller 405 is connected in signal communication with a third input of the deblocking filter 465. A fourth output of the decoder controller 405 is connected in signal communication with a second input of the intra prediction module 460, a first input of the motion compensator and template matching predictor 470, a second input of the reference picture buffer 480, and a third input of the template searcher 444.

An output of the motion compensator and template matching predictor 470 is connected in signal communication with a first input of a switch 497. An output of the intra prediction module 460 is connected in signal communication with a second input of the switch 497. An output of the switch 497 is connected in signal communication with a first non-inverting input of the combiner 425.

An input of the input buffer 410 is available as an input of the decoder 400, for receiving an input bitstream. A first output of the deblocking filter 465 is available as an output of the decoder 400, for outputting an output picture.

As noted above, the present principles are directed to methods and apparatus for template matching prediction (TMP) in video encoding and decoding.

Also as noted above, the prediction performance of template matching prediction is highly dependent on the correlation between the target block and its corresponding template. It is well known in the art to use the reconstructed pixels surrounding the target block, on the top and/or to the left, as shown in FIG. 1. However, utilizing template matching prediction in such a manner lessens encoder and decoder performance, especially when the template has a lower correlation with the targeted signal.

We recognize this limitation and therefore postulate an improved method of template matching prediction for implicit motion derivation. Thus, we describe a new template matching technique for multi-hypothesis inter-prediction, where motion

information for some hypothesis is derived by template matching. In doing so, it is effectively a synergistic combination of template matching prediction and explicit motion prediction. This reduces the bitrate overhead for motion information, thus improving the coding efficiency. It is to be appreciated that while embodiments of the present principles are primarily described herein in the context of inter-prediction, given the teachings of the present principles provided herein, one of ordinary skill in this and related arts will readily understand that the present principles can also be applied for use with intra-prediction.

Thus, in accordance with the present principles, we describe methods and apparatus to solve the limitation of template matching prediction and the limitations of explicit and implicit multiple-hypothesis motion compensation predictions. Also, in particular, we describe using template matching prediction to do joint implicit and explicit prediction to improve multiple-hypothesis motion compensated prediction. In previous solutions, the template is taken from the reconstructed surrounding pixels of the target pixels/block. The surrounding pixels sometimes fail to represent the target pixels, thus the performance of template matching prediction can be lower than the traditional block matching scheme. However, in accordance with one or more embodiments of the present principles, we describe methods and apparatus in which the template can be the prediction of the target pixels/block, as shown in FIG. 5. Turning to FIG. 5, an example of template matching prediction (TMP) scheme combined with single prediction (uni-prediction) is indicated generally by the reference numeral 500. The template matching prediction scheme with single prediction 500 involves a current frame 510, a reference frame in List 0 520, a reference frame in List 1 530. The current frame 510 includes a target block 511. The reference frame in List 0 520 includes a prediction 521 (also designated " P_0 "). The reference frame in List 1 530 includes a prediction 531 (also designated " P_1 "). A motion vector MV_0 is used to denote the motion between the target block 511 and the prediction 521. If the prediction P_0 is very close to the target block, the performance of the template matching prediction can be very efficient by generating a template based on prediction P_0 . In one embodiment, the template can be the inter prediction of the current block. In a second embodiment, the template can be the intra prediction of the current block. We can send some side information to indicate the exact prediction we use, or we can presume some default prediction. For example, in the first embodiment, we can presume that the inter prediction is

always taken as the one derived from a motion vector predictor, or the motion vectors derived from skip or direct mode, or we can explicitly send the motion vector information for the prediction. In the second embodiment, we can always presume DC intra prediction is used or we can send the side information to indicate which
5 intra prediction in the MPEG-4 AVC Standard is used. In a third embodiment, we can fill the template for the target block using the template search from spatial neighboring pixels.

Multi-hypothesis Using Template Matching

10 In this scheme, the template is achieved by inter-prediction by explicit motion estimation. The motion vector can be derived from a motion vector predictor, such as in the MPEG-4 AVC Standard, or can be explicitly sent in the bitstream. Then, we use the template to search the available reference pictures to obtain more motion compensated predictions. In this way, we can take advantage of both explicit motion
15 estimation and implicit motion estimation. Since we can obtain multiple predictions from the same or different reference pictures without incurring more overhead, we can implement any number of hypotheses with this method.

In the following, we shall describe four different embodiments in which template matching prediction is used to improve multi-hypothesis motion
20 compensated prediction. The first two embodiments are focused on B frames. The third embodiment is discussed from the view of multiple reference pictures. The present principles are also applicable to P frames. In the fourth embodiment, we describe how this methodology can also be used for intra picture coding.

25 **Embodiment 1:** TMP combined with single reference list prediction (uni-prediction)

In this embodiment, explicit motion estimation is done for one reference list to obtain the prediction data (P_0) for the target block, which is then used as a template to search in the other reference list to obtain the prediction data (P_1). Then, the final
30 prediction (P) is the weighted average of P_0 and P_1 . The weight can be a simple average (weight equal to 0.5) or other weights based on some distortion measurement, reference distance to the current picture, and so forth. The example is illustrated in FIG. 5.

This approach can be easily combined with a uni-prediction mode in the MPEG-4 AVC Standard such as, for example, but not limited to, mode MxNL0, or MxNL1, where M, N = 16, 8, 4, L0 and L1 represents the prediction coming from List0 or List1 respectively. We can add a flag in the macroblock level or
 5 submacroblock level to indicate whether uni-prediction or bi-prediction with template matching prediction is used for this mode.

bi_tmp_flag equal to 1 specifies that bi-prediction using uni-prediction and template matching prediction is applied. **bi_tmp_flag** equal to 0 specifies that uni-
 10 prediction is applied.

One example encoding procedure for 16x16L0 in accordance with an embodiment is illustrated as follows:

- 15 (1) Perform motion estimation for a macroblock 16x16 in reference List 0 to obtain P_0 , and compute RDcost for P_0 .
- (2) Use P_0 as a template to search in reference List 1 to obtain P_1 , and compute RDcost for $0.5(P_0 + P_1)$.
- 20 (3) Compare RDcost from (1) and (2) and set **bi_tmp_flag**. If the RDcost of (2) is smaller than that of (1), then set **bi_tmp_flag**=1; otherwise, set **bi_tmp_flag**=0.

The corresponding decoding procedure for 16x16L0 in accordance with an embodiment is illustrated as follows:

- 25 (1) Parse and decode the syntax of **bi_tmp_flag**, and motion vectors for 16x16L0.
- (2) Perform motion compensation for a macroblock 16x16 in reference List 0 to obtain P_0 .
- (3) If **bi_tmp_flag** is equal to 0, then stop. Otherwise,
- 30 (4) perform: use P_0 as a template to search in reference List 1 to obtain P_1 , and obtain the final prediction as $0.5(P_0 + P_1)$.

We should note that for template searching the template size could be different from the size of P_0 . For example, we can perform template matching prediction using an 8x8 block size instead of a 16x16 block size for the search of P_1 .

with 16x16L0. Alternatively, we can perform template matching prediction using 20x20 in which we can increase the size of the template by incorporating the surrounding pixels of P_0 .

Turning to FIG. 6, a video encoding method using template matching prediction with single reference list prediction (uni-prediction) is indicated generally by the reference numeral 600. The method 600 includes a start block 603 that passes control to a decision block 606. The decision block 606 determines whether or not the current mode is template matching mode (TMP). If so, then control is passed to a decision block 609. Otherwise, control is passed to a function block 663.

The decision block 609 determines the explicit motion estimation direction (forward L0 or backward L1). If explicit motion estimation direction is backward(BAK), control is passed to a function block 612. Otherwise, control is passed to a function block 627.

The function block 612 performs motion estimation in reference List 1 to obtain prediction P_1 and motion vector Mv , and passes control to a function block 615. The function block 615 performs template matching in reference List 0 based on P_1 to obtain P_0 , and passes control to a function block 618. The function block 618 calculates $P = (P_0 + P_1)/2$, and passes control to a decision block 621. The decision block 621 determines whether or not the rate-distortion cost of P_1 , namely $Rd(P_1)$, is less than the rate-distortion cost of P , namely $Rd(P)$. If so, then control is passed to a function block 624. Otherwise, control is passed to a function block 657.

The function block 624 sets $bi_tmp_flag = 0$, sets Prediction = P_1 , and passes control to a function block 642.

The function block 642 computes the residue, and passes control to a function block 645. The function block 645 computes the rate-distortion of the prediction, namely $Rd(\text{Prediction})$, and passes control to a function block 648. The function block 648 performs mode selection, and passes control to a function block 651. The function block 651 encodes the motion vector Mv and other syntax, and passes control to a function block 654. The function block 654 encodes the residue, and passes control to an end block 699.

The function block 657 sets $bi_tmp_flag = 1$, sets Prediction = P , and passes control to the function block 642.

The function block 627 performs motion estimation in reference List 0 to obtain prediction P_0 and motion vector Mv , and passes control to a function block 630. The function block 630 performs template matching in reference List 1 based on P_0 to obtain P_1 , and passes control to a function block 633. The function block 633 calculates $P = (P_0 + P_1)/2$, and passes control to a decision block 636. The decision block 636 determines whether or not the rate-distortion cost of P_0 , namely $Rd(P_0)$, is less than the rate-distortion cost of P , namely $Rd(P)$. If so, then control is passed to a function block 639. Otherwise, control is passed to a function block 660.

The function block 639 sets $bi_tmp_flag = 0$, sets Prediction = P_0 , and passes control to the function block 642.

The function block 660 sets $bi_tmp_flag = 1$, sets Prediction = P , and passes control to the function block 642.

The function block 663 encodes using a non-TMP mode(s), and passes control to the function block 648.

Turning to FIG. 7, a video decoding method using template matching prediction with single reference list prediction (uni-prediction) is indicated generally by the reference numeral 700. The method 700 includes a start block 703 that passes control to a function block 706. The function block 706 parses the syntax, and passes control to a decision block 709. The decision block 709 determines whether or not the current mode is template matching prediction (TMP) mode. If so, then control is passed to a function block 712. Otherwise, control is passed to a function block 751.

The function block 712 decodes the motion vector Mv , the residue, and bi_tmp_flag , and passes control to a decision block 715. The decision block 715 determines the explicit motion estimation direction. If the explicit motion estimation direction is backward (BAK), control is passed to a function block 718. Otherwise, control is passed to a function block 730.

The function block 718 obtains P_1 in reference List 1 with Mv , and passes control to a decision block 721. The decision block 721 determines whether or not $bi_tmp_flag = 1$. If so, then control is passed to a function block 724. Otherwise, control is passed to a function block 745.

The function block 724 performs template matching in reference List 0 based on P_1 to obtain P_0 , and passes control to a function block 727. The function block 727 calculates Prediction = $(P_0 + P_1)/2$, and passes control to a function block 742.

The function block 742 adds the residue to the prediction, compensates the current block, and passes control to an end block 799.

The function block 745 sets Prediction = P_0 , and passes control to the function block 742.

5 The function block 730 obtains P_0 in reference List 0 with Mv, and passes control to a decision block 733. The decision block 733 determines whether or not $bi_tmp_flag = 1$. If so, then control is passed to a function block 736. Otherwise, control is passed to a function block 743.

10 The function block 736 performs template matching in reference List 1 based on P_0 to obtain P_1 , and passes control to a function block 739. The function block 739 calculates Prediction = $(P_0 + P_1)/2$, and passes control to the function block 742.

The function block 748 sets Prediction = P_0 , and passes control to the function block 742.

15 The function block 751 decodes using a non-TMP mode(s), and passes control to the end block 799.

Embodiment 2: TMP combined with bi-prediction

In one embodiment, the traditional bi-prediction is performed and then for each prediction in one reference list, we can find its corresponding template
 20 matching prediction in the other reference list. The final result can be 3 or 4 hypothesis predictions. We should note that in this approach, when multiple reference pictures are enabled, we do not limit the template matching prediction to the same reference picture as that for the explicit prediction. FIG. 8 shows such an example. Turning to FIG. 8, an example of template matching prediction combined
 25 with bi-prediction is indicated generally by the reference numeral 800. The template matching prediction 800 with bi-prediction involves a current frame 810, a reference frame 1 in List 0 820, a reference frame 0 in List 0 830, and a reference frame in List 1 840. The current frame 810 includes a target block 811, the reference frame 1 in List 0 820 includes a prediction 821 (also designated " P_3 "), the reference frame 0 in
 30 List 0 830 include a prediction 831 (also designated " P_3 ") and a prediction 832 (also designated " P_0 "), and the reference frame in List 1 840 includes a prediction 841 (also designated " P_1 ") and a prediction 842 (also designated " P_2 "). We first perform motion estimation in reference List 0 to obtain P_0 and in reference List 1 to obtain P_1 , then we use P_0 as a template to search in reference List 1 to obtain P_2 , and P_1 as a

template to search in reference List 0 to obtain P_3 . When multiple reference pictures are used, P_3 and P_0 do not need to reside in the same reference picture. A motion vector MV0 is used to denote the motion between the target block 811 and the prediction 832, and a motion vector MV1 is used to denote the motion between the target block 811 and the prediction 841.

This approach can be easily combined with bi-prediction mode for the MPEG-4 AVC Standard such as, for example, but not limited to, mode MxNBI or direct mode, where M, N = 16, 8, 4, and BI represents Bi-prediction. We can add an indicator in the macroblock level or sub-macroblock level to specify how many hypotheses are performed with the use of template matching prediction for this mode.

multi_tmp_mode equal to 0 specifies that traditional bi-prediction is applied. **multi_tmp_mode** equal to 1 specifies that 3 hypotheses are applied, and traditional bi-prediction plus template matching prediction using List 0 prediction as the template and searching in List 1 are performed. **multi_tmp_mode** equal to 2 specifies that 3 hypothesis are applied, and traditional bi-prediction plus template matching prediction using List 1 prediction as the template and searching in List 0 are performed. **multi_tmp_mode** equal to 3 specifies that 4 hypothesis are applied, and traditional bi-prediction plus template matching prediction using List 1 prediction as the template, searching in List 0, template matching prediction using List 0 prediction as the template, and searching in List 1 are performed.

One example encoding procedure for 16x16BI in accordance with an embodiment is illustrated as follows:

- (1) Perform motion estimation for a macroblock 16x16 in reference List 0 to obtain P_0 and in reference List 1 to obtain P_1 , and compute RDcost for $(P_0 + P_1)/2$.
- (2) Use P_0 as a template to search in reference List 1 to obtain P_2 , and compute RDcost for $(P_0 + P_1 + P_2)/3$.
- (3) Use P_1 as a template to search in reference List 0 to obtain P_3 , and compute RDcost for $(P_0 + P_1 + P_3)/3$.
- (4) Compute RDcost for $(P_0 + P_1 + P_2 + P_3)/4$.
- (5) Compare RDcost from (1), (2), (3), and (4).

If RDcost of (1) is the smallest, then set multi_tmp_mode to 0.
 If RDcost of (2) is the smallest, then set multi_tmp_mode to 1.
 If RDcost of (3) is the smallest, then set multi_tmp_mode to 2.
 If RDcost of (4) is the smallest, then set multi_tmp_mode to 3.

5

The corresponding decoding procedure for 16x16BI in accordance with an embodiment is illustrated as follows:

- (1) Parse and decode the syntax of multi_tmp_mode and motion vectors for 16x16BI.
- 10 (2) Perform motion compensation for a macroblock 16x16 in reference List 0 to obtain P_0 and in reference List 1 to obtain P_1 .
- (3) If multi_tmp_mode is equal to 0, then the final prediction is $(P_0 + P_1)/2$. Otherwise,
- (4) if multi_tmp_mode is equal to 1, then use P_0 as a template to search in reference List 1 to obtain P_2 , and obtain the final prediction as $(P_0 + P_1 + P_2)/3$.
- 15 (5) If multi_tmp_mode is equal to 2, then use P_1 as a template to search in reference List 0 to obtain P_3 , and obtain the final prediction as $(P_0 + P_1 + P_3)/3$.
- 20 (6) Otherwise, obtain the final prediction as $(P_0 + P_1 + P_2 + P_3)/4$.

Turning to FIG. 9, a video encoding method using template matching prediction with bi-prediction is indicated generally by the reference numeral 900.

The method 900 includes a start block 903 that passes control to a decision block 906. The decision block 906 determines whether or not the current mode is template matching prediction (TMP) mode. If so, then control is passed to a function block 909. Otherwise, control is passed to a function block 954.

The function block 909 performs motion estimation in reference List 0 to obtain prediction P_0 and motion vector Mv0, and passes control to a function block 912. The function block 912 performs motion estimation in reference List 1 to obtain prediction P_1 and motion vector Mv1, and passes control to a function block 915. The function block 915 performs template matching in reference List 1 based on P_0 to obtain P_2 , and passes control to a function block 918. The function block 918 performs template matching in reference List 0 based on P_1 to obtain P_3 , and

passes control to a function block 921. The function block 921 calculates $Pred0 = (P_0 + P_1)/2$, $Pred1 = (P_0 + P_1 + P_2)/3$, $Pred2 = (P_0 + P_1 + P_3)/3$, and $Pred3 = (P_0 + P_1 + P_2 + P_3)/4$, and passes control to a decision block 924. The decision block 924 determines which of the prediction from among $Pred0$, $Pred1$, $Pred2$, and $Pred3$ provide the least rate-distortion (Rd). If $Pred0$ provides the least Rd, then control is passed to a function block 927. If $Pred1$ provides the least Rd, then control is passed to a function block 930. If $Pred2$ provides the least Rd, then control is passed to a function block 933. If $Pred3$ provides the least Rd, then control is passed to a function block 936.

The function block 927 sets $Multi_TMP_Mode=0$, and passes control to a function block 939.

The function block 930 sets $Multi_TMP_Mode=1$, and passes control to the function block 939.

The function block 933 sets $Multi_TMP_mode=2$, and passes control to the function block 939.

The function block 936 sets $Multi_TMP_mode=3$, and passes control to the function block 939.

The function block 939 computes the residue, and passes control to the function block 942. The function block 942 computes the rate-distortion cost of the prediction, namely $Rd(Prediction)$, and passes control to a function block 945. The function block 945 performs mode selection, and passes control to a function block 948. The function block 948 encodes the motion vectors and other syntax, and passes control to a function block 951. The function block 951 encodes the residue, and passes control to an end block 699.

The function block 954 encodes using a non-TMP mode(s), and passes control to the function block 945.

Turning to FIG. 10, a video decoding method using template matching prediction with bi-prediction is indicated generally by the reference numeral 1000. The method 1000 includes a start block 1003 that passes control to a function block 1006. The function block 1006 parses the syntax, and passes control to a decision block 1009. The decision block 1009 determines whether or not the current mode is a template matching prediction (TMP) mode. If so, then control is passed to a function block 1012. Otherwise, control is passed to a function block 1045.

The function block 1012 performs motion compensation to obtain P_0 in List 0 and P_1 in List 1, and passes control to a decision block 1015. The decision block 1015 determines which Multi_TMP_Mode. If mode 0 is Multi_TMP_Mode, then control is passed to a function block 1018. If mode 1 is Multi_TMP_Mode, then control is passed to a function block 1021. If mode 2 is Multi_TMP_Mode, then control is passed to a function block 1027. If mode 3 is Multi_TMP_Mode, then control is passed to a function block 1033.

The function block 1018 calculates $Pred = (P_0 + P_1)/2$, and passes control to a function block 1042.

10 The function block 1021 performs template matching in reference List 1 based on P_0 to obtain P_2 , and passes control to a function block 1024. The function block 1024 calculates $Pred = (P_0 + P_1 + P_2)/2$, and passes control to the function block 1042.

15 The function block 1027 performs template matching in reference List 0 based on P_1 to obtain P_3 , and passes control to a function block 1030. The function block 1030 calculates $Pred = (P_0 + P_1 + P_3)/2$, and passes control to the function block 1042.

20 The function block 1033 performs template matching in reference List 1 based on P_0 to obtain P_2 , and passes control to a function block 1036. The function block 1036 performs template matching in reference List 0 based on P_1 to obtain P_3 , and passes control to a function block 1039. The function block 1039 calculates $Pred = (P_0 + P_1 + P_2 + P_3)/2$, and passes control to the function block 1042.

The function block 1042 adds the residue to the prediction, compensates the current block, and passes control to an end block 1099.

25 The function block 1045 decodes using a non-TMP mode(s), and passes control to the end block 1099.

In another embodiment, we still use bi_tmp_flag , where bi_tmp_flag equal to 1 specifies the multi-hypothesis motion compensated prediction for maximal (up to 4) different motion compensated predictions.

30 In yet another embodiment, we extend the idea from embodiment one, but we set the template as the average of two predictors $P_2 = (P_0 + P_1)/2$, then we use this template to search the template matching prediction P_3 from the reference pictures in both lists. Finally, we average the template and template matching prediction, $(P_2 + P_3)/2$.

Turning to FIG. 11, another video encoding method using template matching prediction with bi-prediction is indicated generally by the reference numeral 900.

The method 1100 includes a start block 1103 that passes control to a decision block 1106. The decision block 1106 determines whether or not the current mode is
5 template matching prediction (TMP) mode. If so, then control is passed to a function block 1109. Otherwise, control is passed to a function block 1145.

The function block 1109 performs motion estimation in reference List 0 to obtain prediction P_0 and motion vector $Mv0$, and passes control to a function block 1112. The function block 1112 performs motion estimation in reference List 1 to
10 obtain prediction P_1 and motion vector $Mv1$, and passes control to a function block 1115. The function block 1115 sets Template $P_2 = (P_0 + P_1)/2$, performs template matching both reference lists to obtain P_3 , and passes control to a function block 1118. The function block 1118 calculates $Pred0 = (P_0 + P_1)/2$ and $Pred1 = (P_2 + P_3)/2$, and passes control to a decision block 1121. The decision block 1121
15 determines which prediction provides the least rate-distortion (Rd). If $Pred0$ provides the least Rd, then control is passed to a function block 1124. If $Pred1$ provides the least Rd, then control is passed to a function block 1127.

The function block 1124 sets $bi_tmp_flag = 0$, and passes control to a function block 1130.

20 The function block 1130 computes the residue, and passes control to a function block 1133. The function block 1133 computes the rate-distortion for the prediction, namely $Rd(Prediction)$, and passes control to a function block 1136. The function block 1136 performs mode selection, and passes control to a function block 1139. The function block 1139 encodes the motion vectors and other syntax, and
25 passes control to a function block 1142. The function block 1142 encodes the residue, and passes control to an end block 1199.

The function block 1127 sets $bi_tmp_flag = 1$, and passes control to the function block 1130.

30 The function block 1145 encodes using a non-TMP mode(s), and passes control to the function block 1136.

Turning to FIG. 12, another video decoding method using template matching prediction with bi-prediction is indicated generally by the reference numeral 1000. The method includes a start block 1205 that passes control to a function block 1210. The function block 1210 parses syntax, and passes control to a decision block 1215.

The decision block 1215 determines whether or not the current mode is TMP mode. If so, then control is passed to a function block 1220. Otherwise, control is passed to a function block 1250.

The function block 1220 performs motion compensation to obtain P_0 in List 0 and P_1 in List 1, and passes control to a decision block 1225. The decision block 1225 determines whether the current value of `bi_tmp_flag` is equal to 0 or 1. If the current value of `bi_tmp_flag` is equal to 0, then control is passed to a function block 1230. Otherwise, control is passed to a function block 1235.

The function block 1230 calculates $\text{Pred} = (P_0 + P_1)/2$, and passes control to a function block 1245.

The function block 1245 adds the residue to the prediction, compensates the current block, and passes control to an end block 1299.

The function block 1235 sets template $P_2 = (P_0 + P_1)/2$, performs template matching in both reference lists to obtain P_3 , and passes control to a function block 1240. The function block 1240 calculates $\text{Pred} = (P_2 + P_3)/2$, and passes control to the function block 1245.

The function block 1250 decodes using a non-TMP mode(s), and passes control to the end block 1299.

Similarly, we can obtain one or more template matching predictions with this method. Also, in another embodiment, one could obtain the final prediction by the following average: $(P_0 + P_1 + P_3)/3$. An additional embodiment would be to consider the N most similar patches to the P_2 average and the final predictor would be the averages: $(P_2 + P_3 + \dots + P_N)/(N-1)$ or $(P_0 + P_1 + P_3 + \dots + P_N)/N$ among other many possibilities.

We should note in this embodiment, if the template matching prediction is the same as the motion compensated prediction, we do not need to add the template matching prediction anymore, i.e., all multi-hypothesis motion compensated predictions should be different.

Embodiment 3: TMP combined with multiple reference pictures

In the MPEG-4 AVC Standard, multiple reference pictures can be employed to further improve the coding efficiency. Thus, we do not need to limit our invention for different reference lists as described with respect to embodiment 1 and embodiment 2 above. We can obtain a template from one reference picture, and

then perform a template matching prediction search in the same reference picture or other reference pictures that are or not in the same list of pictures. This can allow us to implement our invention with one reference list. It also allows us to implement our invention for any number of multi-hypothesis motion compensated predictions.

5 Turning to FIG. 13, an example of template matching prediction with multiple reference pictures is indicated generally by the reference numeral 1300. The template matching prediction with multiple reference pictures 1300 involves a current frame 1310, a reference frame 1 in List 0 1320, and a reference frame 0 in List 0 1330. The current frame 1310 includes a target block 1311, the reference frame 1 in List 0 1320 includes a prediction 1321 (also designated " P_1 "), and the reference frame 0 in List 0 1330 include a prediction 1331 (also designated " P_0 "). A motion vector 1340 is used to denote the motion between the target block 1311 and the prediction 1331. We first perform motion estimation to obtain P_0 in ref frame 0 in reference List 0. We then use P_0 as a template to search in other reference pictures (ref frame 1) of reference List 0 to obtain P_1 .

This approach can be easily combined with prediction mode in P pictures of the MPEG-4 AVC Standard when multiple reference pictures are used. For example, the approach can be used for mode MxN or skip mode, where M or N can be equal to any of 16, 8, and 4. We can add a flag at the macroblock level or sub-macroblock level to indicate how many hypotheses are performed with the use of template matching prediction for this mode.

bi_tmp_flag equal to 1 specifies that bi-prediction using uni-prediction and template matching prediction is applied. **bi_tmp_flag** equal to 0 specifies that uni-prediction is applied.

One example encoding procedure for 16x16 in accordance with an embodiment is illustrated as follows:

- 30 (1) Perform motion estimation for a macroblock 16x16 in reference List 0 to obtain P_0 , and compute RDcost for P_0 .
- (2) Use P_0 as a template to search in other reference pictures of reference List 0 to obtain P_1 , and compute the RDcost for $(P_0 + P_1)/2$.

- (3) Compare the RDcost from (1) and (2) and set bi_tmp_flag . If the RDcost of (2) is smaller than that of (1), then set $bi_tmp_flag=1$; otherwise, set $bi_tmp_flag=0$.

5 The corresponding decoding procedure for 16x16 in accordance with an embodiment is illustrated as follows:

- (1) Parse and decode the syntax of bi_tmp_flag , and motion vectors for 16x16.
- (2) Perform motion compensation for a macroblock 16x16 in reference List 0 to obtain P_0 .
- 10 (3) If bi_tmp_flag is equal to 0, then stop. Otherwise,
- (4) perform: use P_0 as a template to search in other reference pictures of reference List 0 obtain P_1 , and obtain the final prediction as $0.5(P_0 + P_1)$.

15

Turning to FIG. 14, a video encoding method using template matching prediction with multiple reference pictures is indicated generally by the reference numeral 1400. The method 1400 includes a start block 1405 that passes control to a decision block 1410. The decision block 1410 determines whether or not the current mode is TMP mode. If so, then control is passed to a function block 1415. Otherwise, control is passed to a function block 1470.

The function block 1415 performs motion estimation in reference List 0 to obtain prediction P_0 and motion vector Mv , and passes control to a function block 1420. The function block 1420 performs template matching in other reference pictures of List 0 based on P_0 to obtain P_1 , and passes control to a function block 1425. The function block 1425 calculates $P = (P_0 + P_1)/2$, and passes control to a decision block 1430. The decision block 1430 determines whether the rate distortion cost of Prediction P_0 is less than the rate distortion cost of Prediction P , namely $Rd(P_0) < Rd(P)$. If so, then control is passed to a function block 1435. Otherwise, control is passed to a function block 1465.

30 The function block 1435 sets $bi_tmp_flag = 0$, sets Prediction = P_0 , and passes control to a function block 1440.

The function block 1440 computes the residue, and passes control to a function block 1445. The function block 1445 computes $RD(\text{Prediction})$, and passes

control to a function block 1450. The function block 1450 performs mode selection, and passes control to a function block 1455. The function block 1455 encodes the motion vector Mv and other syntax, and passes control to a function block 1460. The function block 1460 encodes the residue, and passes control to an end block 1499.

The function block 1465 sets $bi_tmp_flag = 1$, sets $Prediction = P$, and passes control to the function block 1440.

The function block 1470 encodes using a non-TMP mode(s), and passes control to the function block 1450.

Turning to FIG. 15, a video decoding method using template matching prediction with multiple reference pictures is indicated generally by the reference numeral 1500. The method 1500 includes a start block 1505 that passes control to a function block 1510. The function block 1510 parses syntax, and passes control to a decision block 1515. The decision block 1515 determines whether or not the current mode is TMP mode. If so, then control is passed to a function block 1420. Otherwise, control is passed to a function block 1555.

The function block 1520 decodes the motion vector Mv , the residue, and bi_tmp_flag , and passes control to a function block 1525. The function block 1525 obtains P_0 in reference List 0 with motion vector Mv , and passes control to a decision block 1530. The decision block 1530 determines whether or not $B-TMP_Flag$ is equal to 1. If so, then control is passed to a function block 1535. Otherwise, control is passed to a function block 1550.

The function block 1535 performs template matching in other reference pictures of List 0 based on P_0 to obtain P_1 , and passes control to a function block 1540. The function block 1540 calculates $Prediction = (P_0 + P_1)/2$, and passes control to a function block 1545.

The function block 1545 adds the residue to the prediction, compensates the current block, and passes control to an end block 1599.

The function block 1550 calculates $Prediction = P_0$, and passes control to the function block 1545.

The function block 1555 decodes using a non-TMP mode(s), and passes control to the end block 1599.

Embodiment 4: TMP combined with directional intra prediction or displaced intra prediction for intra picture coding.

In this embodiment, for intra picture coding, we can fill the template for the target block using directional intra prediction in the MPEG-4 AVC Standard, or using displaced intra prediction (DIP). Then, we search the template from the reconstructed pixels in the current picture. We then apply multiple hypothesis prediction by combining template searched predictions and directional/displaced intra prediction.

Turning to FIG. 16, another video encoding method using template matching prediction for intra prediction is indicated generally by the reference numeral 1600. The method 1600 includes a start block 1605 that passes control to a decision block 1610. The decision block 1610 determines whether or not the current mode is TMP mode. If so, then control is passed to a decision block 1615. Otherwise, control is passed to a function block 1675.

The decision block 1615 determines whether or not directional/displaced intra prediction (DIP) is allowed. If so, then control is passed to a function block 1620. Otherwise, control is passed to a function block 1680.

The function block 1620 performs intra motion estimation to obtain P_0 , and passes control to a function block 1625.

The function block 1625 performs intra template matching (i.e., search the template from the reconstructed area of the current frame) based on P_0 to obtain P_1 , and passes control to a function block 1630. The function block 1630 calculates $P = (P_0 + P_1)/2$, and passes control to a decision block 1635. The decision block 1635 determines whether or not the rate distortion cost of Prediction P_0 is less than the rate distortion cost of Prediction P , namely $Rd(P_0) < Rd(P)$. If so, then control is passed to a function block 1640. Otherwise, control is passed to a function block 1670.

The function block 1640 sets $bi_tmp_flag=0$, sets $Pred = P_0$, and passes control to a function block 1645.

The function block 1645 computes the residue, and passes control to a function block 1650. The function block 1650 computes $Rd(Pred)$, and passes control to a function block 1655. The function block 1655 performs mode selection, and passes control to a function block 1660. The function block 1660 encodes the prediction mode, motion vector, and other syntax, and passes control to a function

block 1665. The function block 1665 encodes the residue, and passes control to an end block 1699.

The function block 1670 sets $bi_tmp_flag = 1$, sets $Pred = P$, and passes control to the function block 1645.

5 The function block 1675 encodes using a non-TMP mode(s), and passes control to the function block 1655.

Turning to FIG. 17, another video decoding method using template matching prediction with multiple reference pictures is indicated generally by the reference numeral 1700. The method 1700 includes a start block 1705 that passes control to a function block 1710. The function block 1710 parses syntax, and passes control to a decision block 1715. The decision block 1715 determines whether or not the current mode is TMP mode. If so, then control is passed to a decision block 1720. Otherwise, control is passed to a function block 1760.

10 The decision block 1720 determines whether or not directional/displaced intra prediction (DIP) is enabled. If so, then control is passed to a function block 1725. Otherwise, control is passed to a function block 1750.

The function block 1725 performs intra motion compensation to obtain P_0 , and passes control to a decision block 1730.

15 The decision block 1730 determines whether or not $bi_tmp_flag=1$. If so, then control is passed to a function block 1735. Otherwise, control is passed to a function block 1755.

The function block 1735 performs intra template matching based on P_0 to obtain P_1 , and passes control to a function block 1740. The function block 1740 calculates $P = (P_0 + P_1)/2$, and passes control to a function block 1745.

20 The function block 1745 adds the residue to P , compensates the current block, and passes control to an end block 1799.

The function block 1750 performs intra prediction to obtain P_0 , and passes control to the decision block 1730.

25 The function block 1755 sets $P = P_0$, and passes control to the function block 1745.

30 The function block 1760 decodes using a non-TMP intra mode(s), and passes control to the end block 1799.

Template Search Simplification

In accordance with an embodiment the present principles, the template search is performed at both an encoder and a decoder. The straightforward method is to use a full (brut force) search. However, the computational complexity for this method is quite significant. In order to reduce the complexity, we describe and propose several methods to simplify the searching algorithm. In one such embodiment, we use a fast search algorithm such as, for example, a diamond search, a logarithm search, an enhanced predictive zonal search (EPZS), and so forth. In a second of such embodiments, presuming that the motion is linear, we can derive the template search predictor from the motion vector that the template used, and scale the template search predictor according to the distance and direction from reference picture to target picture. We then use this predictor as the center and do a small refinement.

Impact On Other Coding Blocks

Motion Vector for TMP

In the MPEG-4 AVC Standard, the motion vector for the current block is predicted from the neighboring block. Thus, the value of the motion vector of the current block will affect the future neighboring blocks. This raises a question regarding the template searching block as to what motion vector we should use, since we have the explicit motion vector and template searched motion vectors. In the first embodiment, after the template search is done, we can retrieve the explicit motion vector for the template. This motion vector is used for future motion vector prediction. In the second embodiment, we presume the motion vector to be some predictor, such as 0, or the median predictor of the current block.

Deblocking Filter

For the deblocking filter, in the first embodiment, we can treat the template matching prediction block the same as an explicit motion estimation block, and use the motion vector for template matching prediction as described above. Then, the deblocking process is not

changed. In the second embodiment, since the template matching prediction has different characteristic than the explicit motion estimation block, we can adjust the block strength, the filter type, and/or filter length accordingly.

5

Independent Parsing

To separate parsing from the decoding process, we cannot perform the template search in the parsing. Therefore, in the parsing process, for motion vector derivation for template matching prediction, if embodiment 1 of motion vector for TMP (as described above) is used, then we can use embodiment 2 for parsing that involves setting the second motion vector by its motion vector predictor, and use embodiment 1 for real decoding.

10

15

TABLE 1 shows exemplary slice header syntax, in accordance with an embodiment of the present principles. TABLE 2 shows exemplary macroblock layer syntax, in accordance with an embodiment of the present principles.

TABLE 1

slice_header() {		
first_mb_in_slice	2	ue(v)
slice_type	2	ue(v)
pic_parameter_set_id	2	ue(v)
...		
if (slice_type != I)		
tmp_enable_flag	2	u(1)
...		

20

TABLE 2

macroblock_layer() {		
mb_type	2	ue(v) ae(v)
if(MbPartPredMode(mb_type, 0) != Intra_4x4 && MbPartPredMode(mb_type, 0) != Intra_8x8 && MbPartPredMode(mb_type, 0) != Intra_16x16)		
bi_tmp_flag	2	u(1)
.....		

The semantics of some of the syntax elements from TABLES 1 and 2 will now be described in further detail as follows:

25

tmp_enable_flag equal to 1 specifies that template matching prediction is enabled for the slice. **tmp_enable_flag** equal to 0 specifies that template matching prediction is not enabled for the slice.

5

bi_tmp_flag equal to 0 specifies template matching prediction is not enabled for the macroblock. **bi_tmp_flag** equal to 1 specifies template matching prediction is enabled for the macroblock. When **bi_tmp_flag** is equal to 1, uni-prediction and template matching prediction are applied for single (uni) prediction modes, and multi-hypothesis motion compensation prediction for maximal (up to 4) different motion compensated predictions is applied for bi-prediction modes.

10

A description will now be given of some of the many attendant advantages/features of the present invention, some of which have been mentioned above. For example, one advantage/feature is an apparatus having an encoder for encoding a target block in a picture by determining at least one predictor for the target block and respectively utilizing the at least one predictor as a template to search for at least one additional predictor for the target block.

15

Another advantage/feature is the apparatus having the encoder as described above, wherein the at least one predictor is an intra predictor, an inter predictor, or a combination thereof.

20

Yet another advantage/feature is the apparatus having the encoder wherein the at least one predictor is an intra predictor, an inter predictor, or a combination thereof as described above, wherein side information for the at least one predictor is at least one of implicitly coded and explicitly coded.

25

Still another advantage/feature is the apparatus having the encoder as described above, wherein the at least one additional predictor is obtained using a template matching prediction process and a multi-hypothesis motion compensation prediction process.

30

Moreover, another advantage/feature is the apparatus having the encoder wherein the at least one additional predictor is obtained using a template matching prediction process and a multi-hypothesis motion compensation prediction process as described above, wherein an explicit motion estimation process is used to determine the at least one predictor and the template matching prediction process is

used to search for the at least one additional predictor with respect to multiple reference lists.

Further, another advantage/feature is the apparatus having the encoder wherein the explicit motion estimation process and the template matching prediction
5 process are used as described above, wherein the explicit motion estimation process is constrained to a particular one of the multiple reference lists, and the template matching prediction process is constrained to a different one of the multiple reference lists.

Also, another advantage/feature is the apparatus having the encoder wherein
10 the explicit motion estimation process and the template matching prediction process are used as described above, wherein the at least one predictor includes a first predictor and a second predictor obtained by respectively applying the explicit motion estimation process and an implicit motion estimation process with respect to
15 a first reference list and a second reference list from among the multiple reference lists, and wherein the at least one additional predictor includes a first additional predictor and a second additional predictor obtained by respectively utilizing the first predictor and the second predictor in the template matching prediction process with respect to the second reference list and the first reference list.

Additionally, another advantage/feature is the apparatus having the encoder
20 wherein the at least one predictor includes a first predictor and a second predictor and the at least one additional predictor includes a first additional predictor and a second additional predictor as described above, wherein the first predictor and the second additional predictor reside in different reference pictures.

Moreover, another advantage/feature is the apparatus having the encoder
25 wherein the explicit motion estimation process and the template matching prediction process are used as described above, wherein the explicit motion estimation process is used in a particular one reference picture corresponding to a particular one of the multiple reference lists, and the template matching prediction process is used in the particular one reference picture or a different reference picture
30 corresponding to a different one of the multiple reference lists.

Further, another advantage/feature is the apparatus having the encoder wherein the at least one additional predictor is obtained using a template matching prediction process and a multi-hypothesis motion compensation prediction process

as described above, wherein the at least one prediction is derived or estimated using a motion vector predictor.

These and other features and advantages of the present principles may be readily ascertained by one of ordinary skill in the pertinent art based on the teachings herein. It is to be understood that the teachings of the present principles may be implemented in various forms of hardware, software, firmware, special purpose processors, or combinations thereof.

Most preferably, the teachings of the present principles are implemented as a combination of hardware and software. Moreover, the software may be implemented as an application program tangibly embodied on a program storage unit. The application program may be uploaded to, and executed by, a machine comprising any suitable architecture. Preferably, the machine is implemented on a computer platform having hardware such as one or more central processing units ("CPU"), a random access memory ("RAM"), and input/output ("I/O") interfaces. The computer platform may also include an operating system and microinstruction code. The various processes and functions described herein may be either part of the microinstruction code or part of the application program, or any combination thereof, which may be executed by a CPU. In addition, various other peripheral units may be connected to the computer platform such as an additional data storage unit and a printing unit.

It is to be further understood that, because some of the constituent system components and methods depicted in the accompanying drawings are preferably implemented in software, the actual connections between the system components or the process function blocks may differ depending upon the manner in which the present principles are programmed. Given the teachings herein, one of ordinary skill in the pertinent art will be able to contemplate these and similar implementations or configurations of the present principles.

Although the illustrative embodiments have been described herein with reference to the accompanying drawings, it is to be understood that the present principles is not limited to those precise embodiments, and that various changes and modifications may be effected therein by one of ordinary skill in the pertinent art without departing from the scope or spirit of the present principles. All such changes and modifications are intended to be included within the scope of the present principles as set forth in the appended claims.

CLAIMS:

1. An apparatus, comprising:
- 5 an encoder (300) for encoding a target block in a picture by determining at least one predictor for the target block and respectively utilizing the at least one predictor as a template to search for at least one additional predictor for the target block.
- 10 2. The apparatus of claim 1, wherein the at least one predictor is an intra predictor, an inter predictor, or a combination thereof.
3. The apparatus of claim 2, wherein side information for the at least one predictor is at least one of implicitly coded and explicitly coded.
- 15 4. The apparatus of claim 1, wherein the at least one additional predictor is obtained using a template matching prediction process and a multi-hypothesis motion compensation prediction process.
- 20 5. The apparatus of claim 4, wherein an explicit motion estimation process is used to determine the at least one predictor and the template matching prediction process is used to search for the at least one additional predictor with respect to multiple reference lists.
- 25 6. The apparatus of claim 5, wherein the explicit motion estimation process is constrained to a particular one of the multiple reference lists, and the template matching prediction process is constrained to a different one of the multiple reference lists.
- 30 7. The apparatus of claim 5, wherein the at least one predictor comprises a first predictor and a second predictor obtained by respectively applying the explicit motion estimation process and an implicit motion estimation process with respect to a first reference list and a second reference list from among the multiple reference lists, and wherein the at least one additional predictor comprises a first additional

predictor and a second additional predictor obtained by respectively utilizing the first predictor and the second predictor in the template matching prediction process with respect to the second reference list and the first reference list.

5 8. The apparatus of claim 7, wherein the first predictor and the second additional predictor reside in different reference pictures.

 9. The apparatus of claim 5, wherein the explicit motion estimation
10 process is used in a particular one reference picture corresponding to a particular
one of the multiple reference lists, and the template matching prediction process is
used in the particular one reference picture or a different reference picture
corresponding to a different one of the multiple reference lists.

 10. The apparatus of claim 4, wherein the at least one prediction is derived
15 or estimated using a motion vector predictor.

 11. A method, comprising:
 encoding a target block in a picture by determining at least one predictor for
the target block and respectively utilizing the at least one predictor as a template to
20 search (615, 630, 915, 918, 1115, 1420, 1625) for at least one additional predictor
for the target block.

 12. The method of claim 11, wherein the at least one predictor is an intra
25 predictor (1620, 1680), an inter predictor (612, 627, 909, 912, 1109, 1112, 1415), or
a combination thereof (1115).

 13. The method of claim 12, wherein side information for the at least one
predictor is at least one of implicitly coded and explicitly coded (651, 948, 1139,
1455, 1660).

30 14. The method of claim 11, wherein the at least one additional predictor is
obtained using a template matching prediction process and a multi-hypothesis
motion compensation prediction process (618, 633, 921, 1118, 1425, 1630).

15. The method of claim 14, wherein an explicit motion estimation process (612, 627, 909, 912, 1109, 1112, 1415) is used to determine the at least one predictor and the template matching prediction process (615, 630, 915, 918, 1115, 1420, 1625) is used to search for the at least one additional predictor with respect to multiple reference lists.

16. The method of claim 15, wherein the explicit motion estimation process (612, 627, 909, 912, 1109, 1112, 1415) is constrained to a particular one of the multiple reference lists, and the template matching prediction process (615, 630, 915, 918, 1115, 1625) is constrained to a different one of the multiple reference lists.

17. The method of claim 15, wherein the at least one predictor comprises a first predictor and a second predictor obtained by respectively applying the explicit motion estimation process (612, 627, 909, 912, 1109, 1112, 1415) and an implicit motion estimation process (615, 630, 915, 918, 1115, 1625) with respect to a first reference list and a second reference list from among the multiple reference lists, and wherein the at least one additional predictor comprises a first additional predictor and a second additional predictor obtained by respectively utilizing the first predictor and the second predictor in the template matching prediction process (615, 630, 915, 918, 1115, 1625) with respect to the second reference list and the first reference list.

18. The method of claim 17, wherein the first predictor and the second additional predictor reside in different reference pictures (1420).

19. The method of claim 15, wherein the explicit motion estimation process (612, 627, 909, 912, 1109, 1112, 1415) is used in a particular one reference picture corresponding to a particular one of the multiple reference lists, and the template matching prediction process (615, 630, 915, 918, 1115, 1420, 1625) is used in the particular one reference picture or a different reference picture corresponding to a different one of the multiple reference lists.

20. The method of claim 14, wherein the at least one prediction is derived or estimated using a motion vector predictor.

21. An apparatus, comprising:

5 a decoder (400) for decoding a target block in a picture by determining at least one predictor for the target block and respectively utilizing the at least one predictor as a template to search for at least one additional predictor for the target block.

10 22. The apparatus of claim 21, wherein the at least one predictor is an intra predictor, an inter predictor, or a combination thereof.

23. The apparatus of claim 22, wherein side information for the at least one predictor is at least one of derived and explicitly decoded.

15

24. The apparatus of claim 21, wherein the at least one additional predictor is obtained using a template matching prediction process and a multi-hypothesis motion compensation prediction process.

20 25. The apparatus of claim 24, wherein an explicit motion estimation process is used to determine the at least one predictor and the template matching prediction process is used to search for the at least one additional predictor with respect to multiple reference lists.

25 26. The apparatus of claim 25, wherein the explicit motion estimation process is constrained to a particular one of the multiple reference lists, and the template matching prediction process is constrained to a different one of the multiple reference lists.

30 27. The apparatus of claim 25, wherein the at least one predictor comprises a first predictor and a second predictor obtained by respectively applying the explicit motion estimation process and an implicit motion estimation process with respect to a first reference list and a second reference list from among the multiple reference lists, and wherein the at least one additional predictor comprises a first

additional predictor and a second additional predictor obtained by respectively utilizing the first predictor and the second predictor in the template matching prediction process with respect to the second reference list and the first reference list.

5

28. The apparatus of claim 27, wherein the first predictor and the second additional predictor reside in different reference pictures.

29. The apparatus of claim 25, wherein the explicit motion estimation
10 process is used in a particular one reference picture corresponding to a particular one of the multiple reference lists, and the template matching prediction process is used in the particular one reference picture or a different reference picture corresponding to a different one of the multiple reference lists.

15 30. The apparatus of claim 24, wherein the at least one prediction is derived or estimated using a motion vector predictor.

31. A method, comprising:

20 decoding a target block in a picture by determining at least one predictor for the target block and respectively utilizing the at least one predictor as a template to search for at least one additional predictor for the target block (724, 736, 1021, 1027, 1235, 1535, 1735).

25 32. The method of claim 31, wherein the at least one predictor is an intra predictor (1725, 1750), an inter predictor (718, 730, 1012, 1220, 1525), or a combination thereof.

30 33. The method of claim 32, wherein side information for the at least one predictor is at least one of derived and explicitly decoded (709, 712, 721, 733, 1009, 1015, 1215, 1225, 1515, 1520, 1715, 1730).

34. The method of claim 31, wherein the at least one additional predictor is obtained using a template matching prediction process (724, 736, 1021, 1027, 1235,

1535, 1735) and a multi-hypothesis motion compensation prediction process (727, 739, 1018, 1024, 1030, 1039, 1230, 1240, 1540, 1740).

5 35. The method of claim 34, wherein an explicit motion estimation process (718, 730, 1012, 1220, 1525) is used to determine the at least one predictor and the template matching prediction process (724, 736, 1021, 1027, 1235, 1735) is used to search for the at least one additional predictor with respect to multiple reference lists.

10 36. The method of claim 35, wherein the explicit motion estimation process (718, 730, 1012, 1220, 1525) is constrained to a particular one of the multiple reference lists, and the template matching prediction process (724, 736, 1021, 1027, 1235, 1735) is constrained to a different one of the multiple reference lists.

15 37. The method of claim 35, wherein the at least one predictor comprises a first predictor and a second predictor obtained by respectively applying the explicit motion estimation process (718, 730, 1012, 1220, 1525) and an implicit motion estimation process (724, 736, 1021, 1027, 1235, 1735) with respect to a first reference list and a second reference list from among the multiple reference lists,
20 and wherein the at least one additional predictor comprises a first additional predictor and a second additional predictor obtained by respectively utilizing the first predictor and the second predictor in the template matching prediction process (724, 736, 1021, 1027, 1235, 1735) with respect to the second reference list and the first reference list.

25 38. The method of claim 37, wherein the first predictor and the second additional predictor reside in different reference pictures (1535).

30 39. The method of claim 35, wherein the explicit motion estimation process (718, 730, 1012, 1220, 1525) is used in a particular one reference picture corresponding to a particular one of the multiple reference lists, and the template matching prediction process (724, 736, 1021, 1027, 1235, 1535, 1735) is used in the particular one reference picture or a different reference picture corresponding to a different one of the multiple reference lists.

40. The method of claim 34, wherein the at least one prediction is derived or estimated using a motion vector predictor.

5 41. A storage media having video signal data encoded thereupon, comprising:

a target block in a picture encoded by determining at least one predictor for the target block and respectively utilizing the at least one predictor as a template to search for at least one additional predictor for the target block.

10

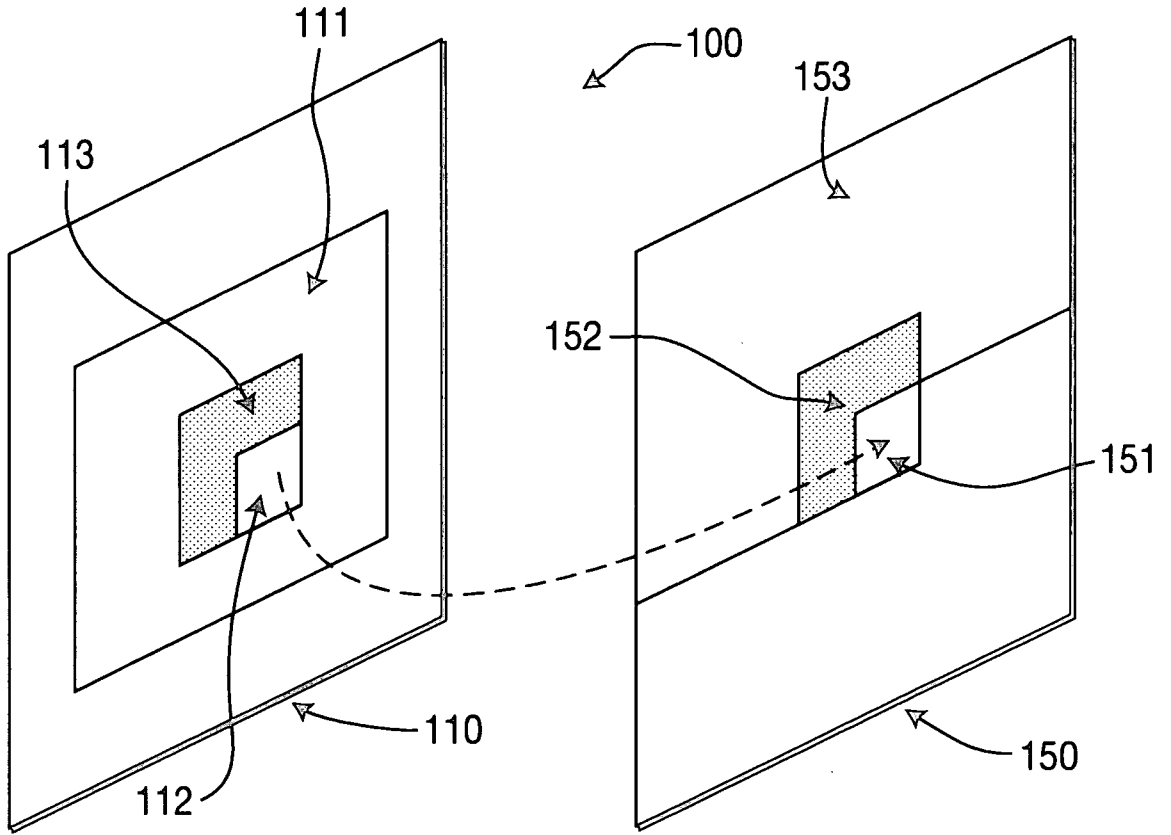


FIG. 1

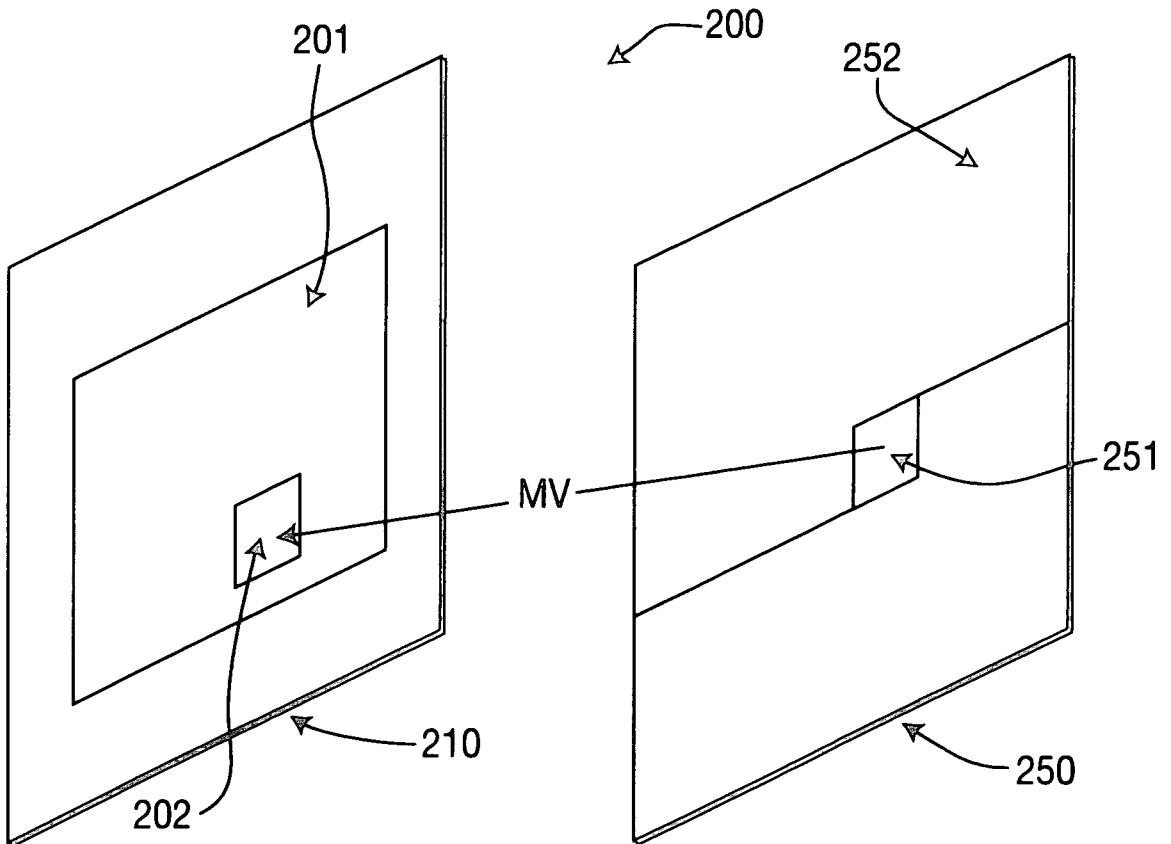


FIG. 2

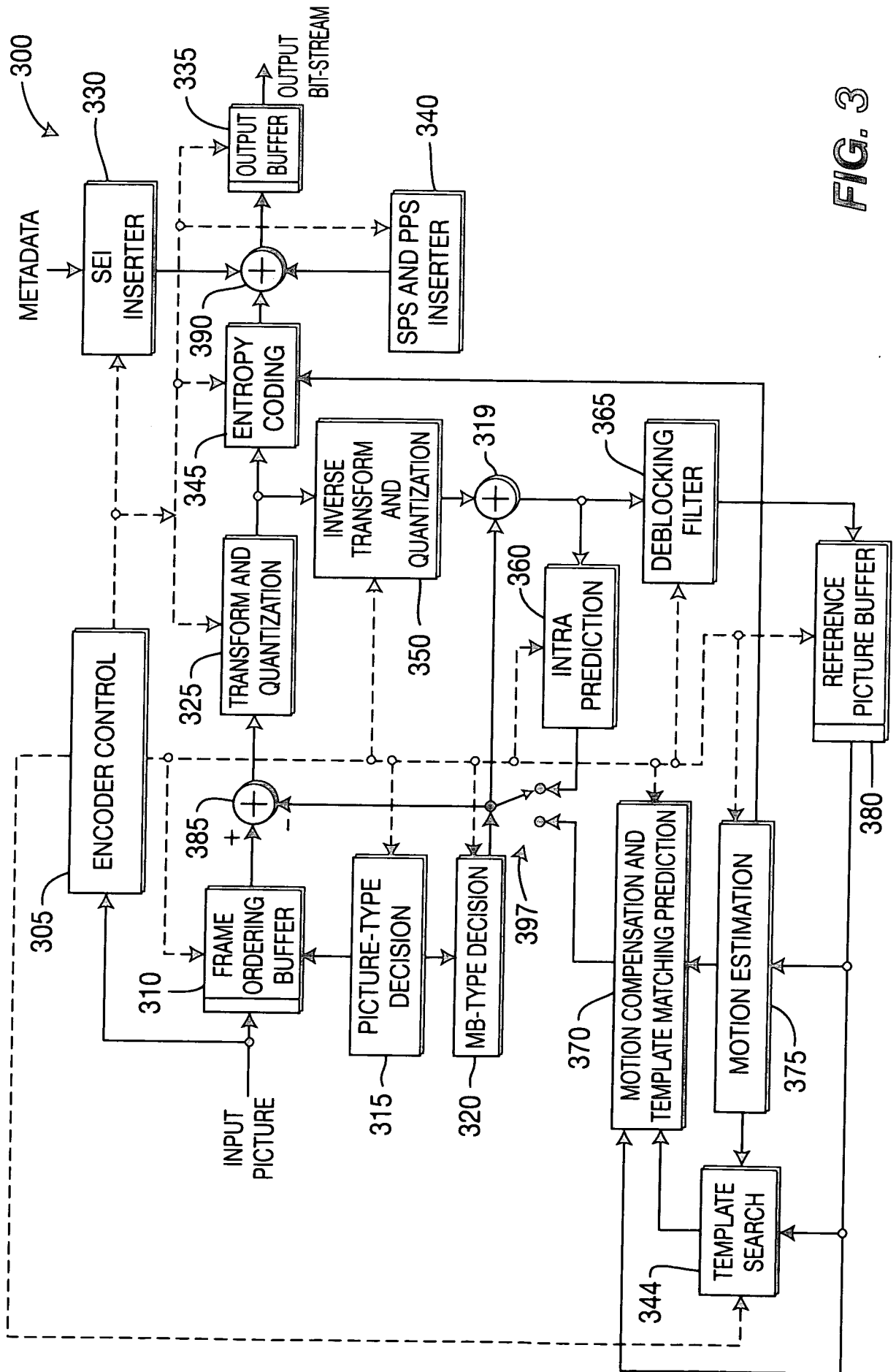


FIG. 3

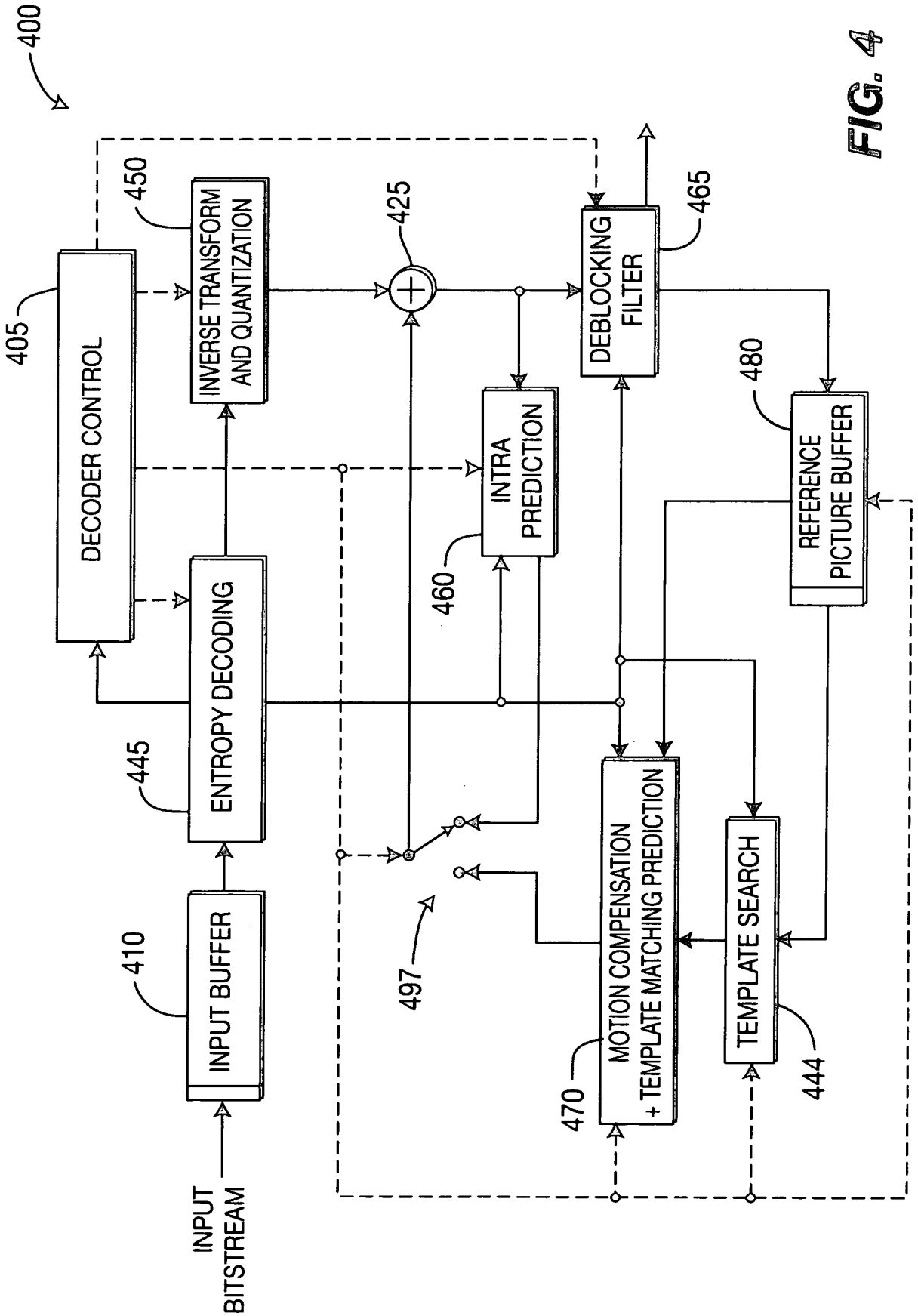


FIG. 4

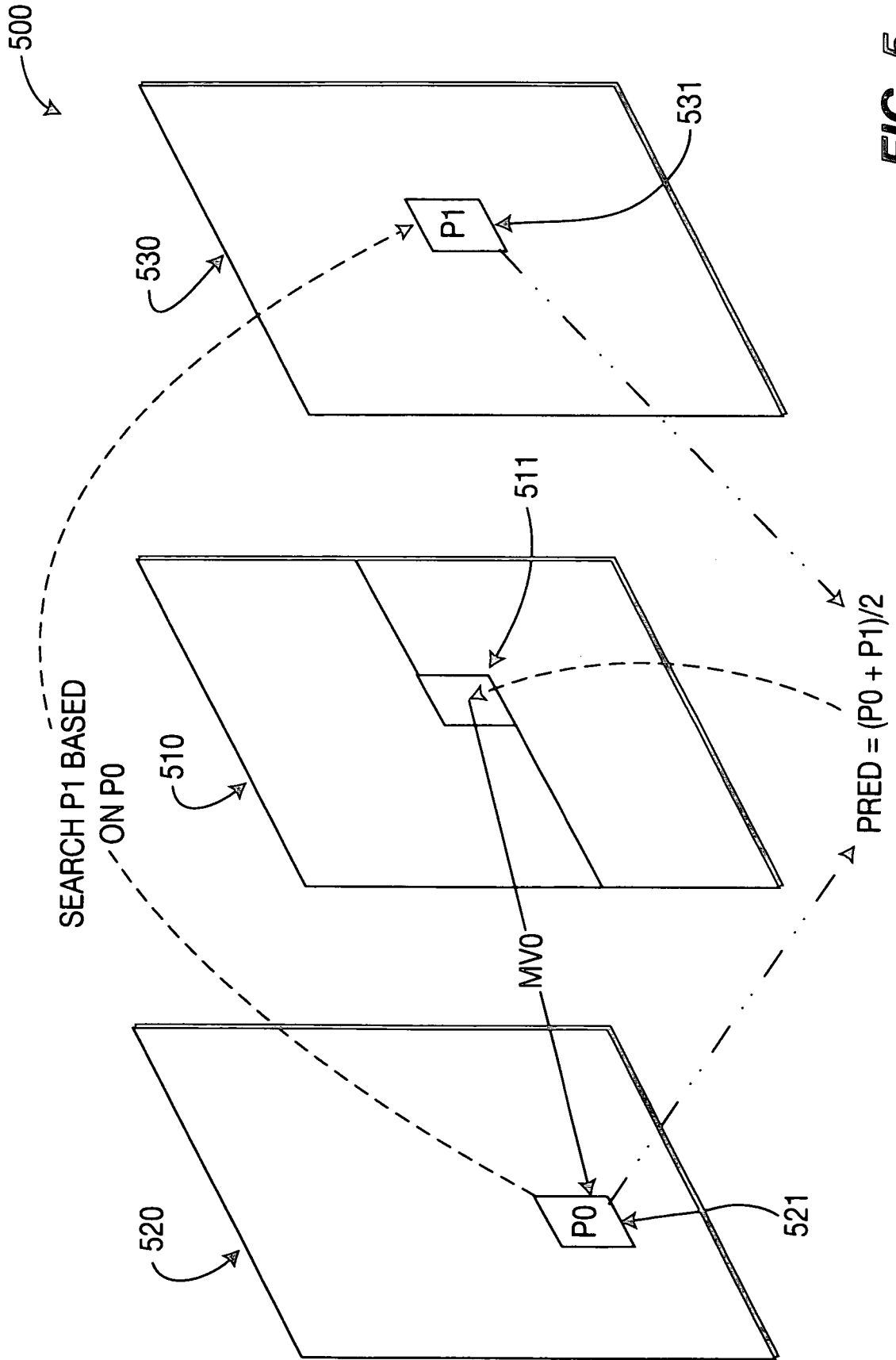


FIG. 5

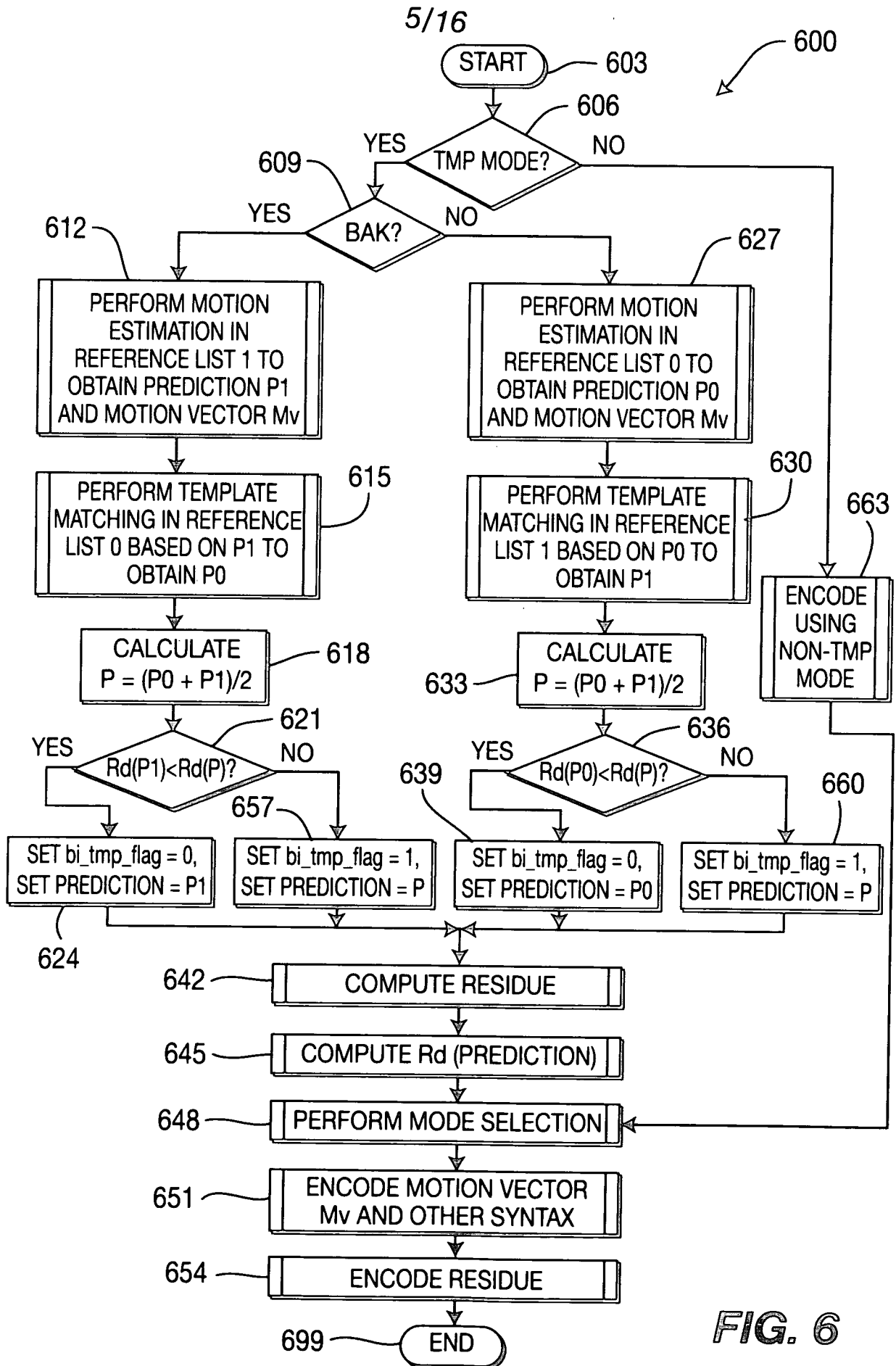


FIG. 6

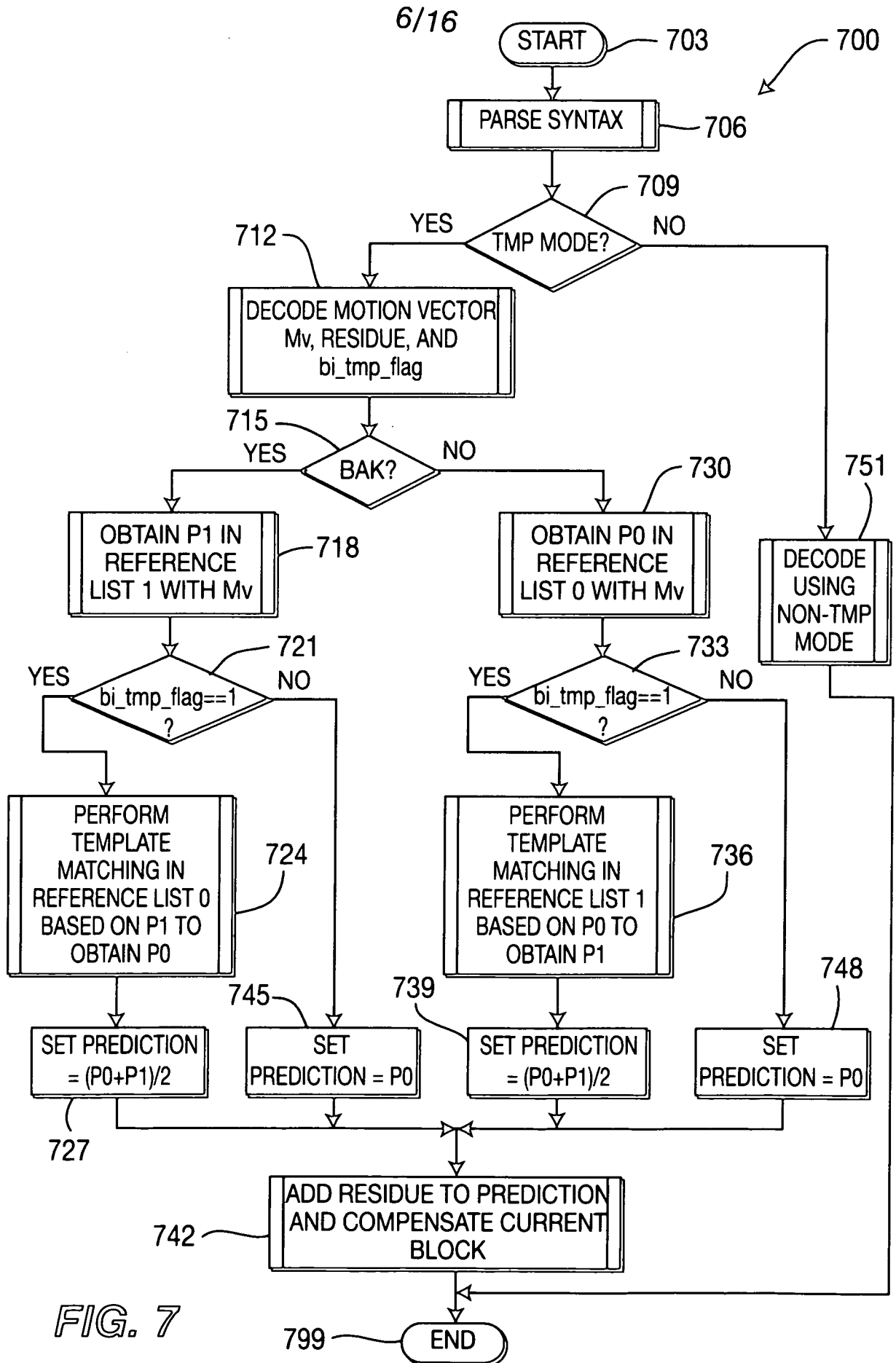


FIG. 7

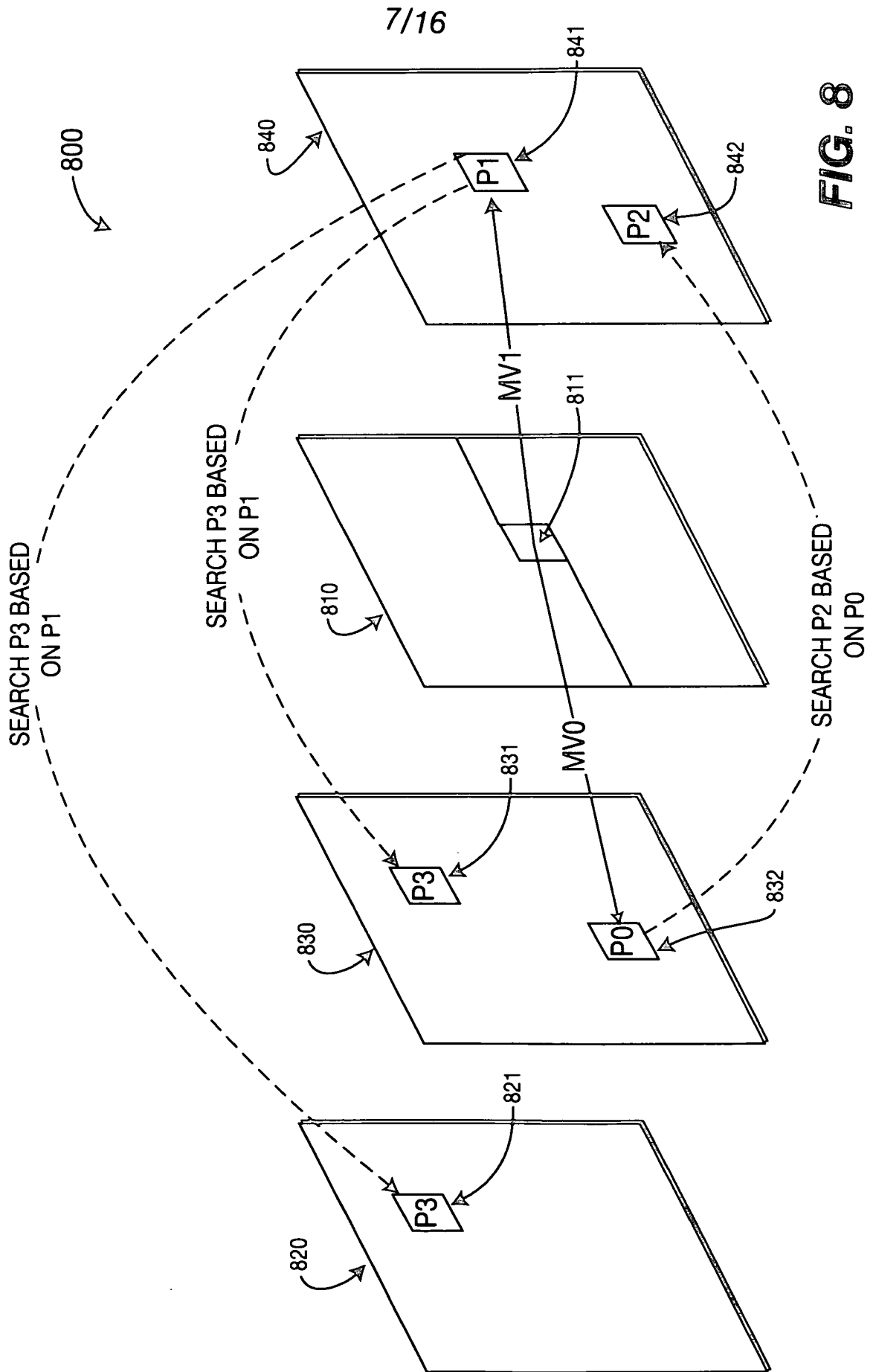


FIG. 8

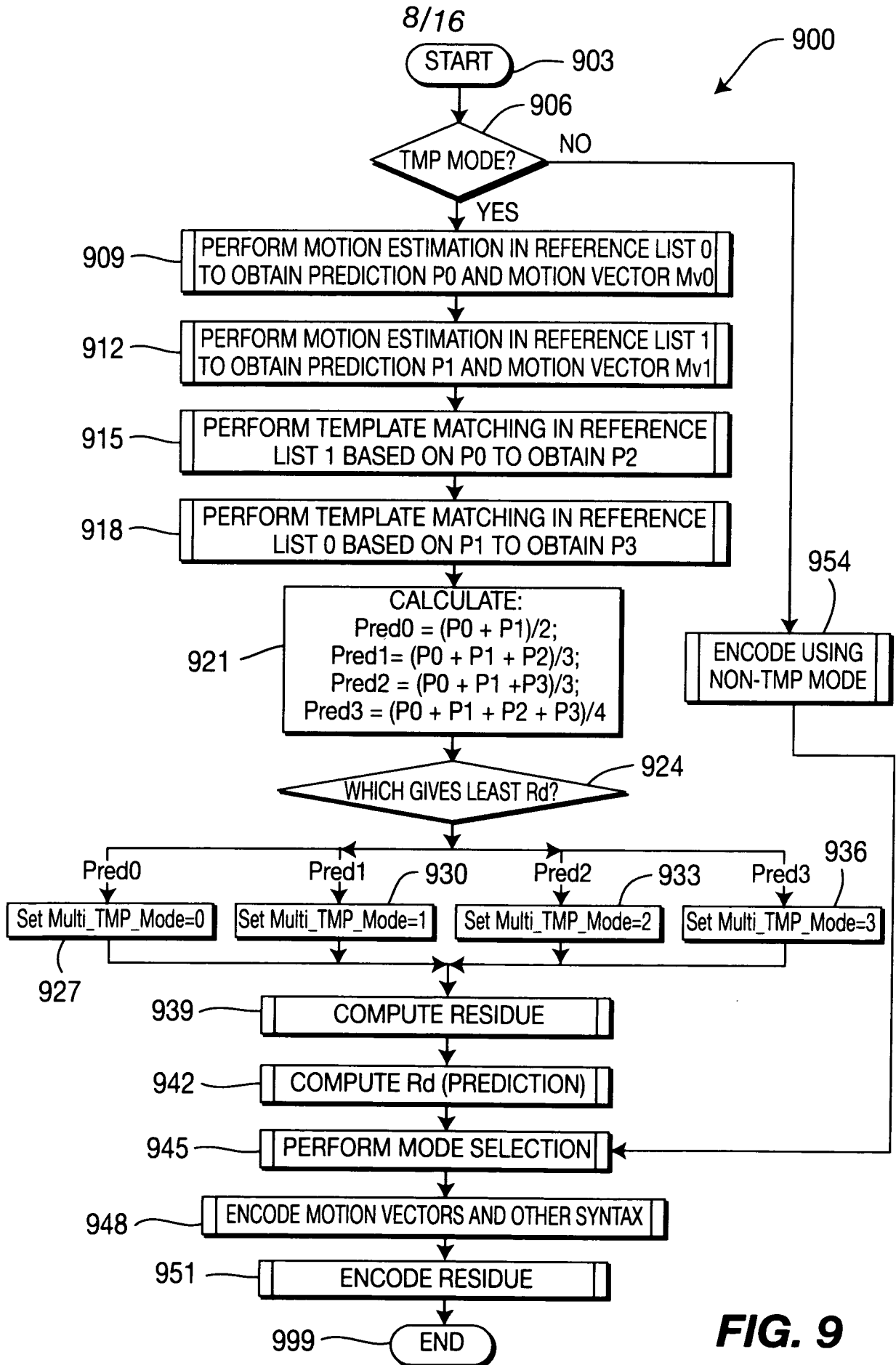
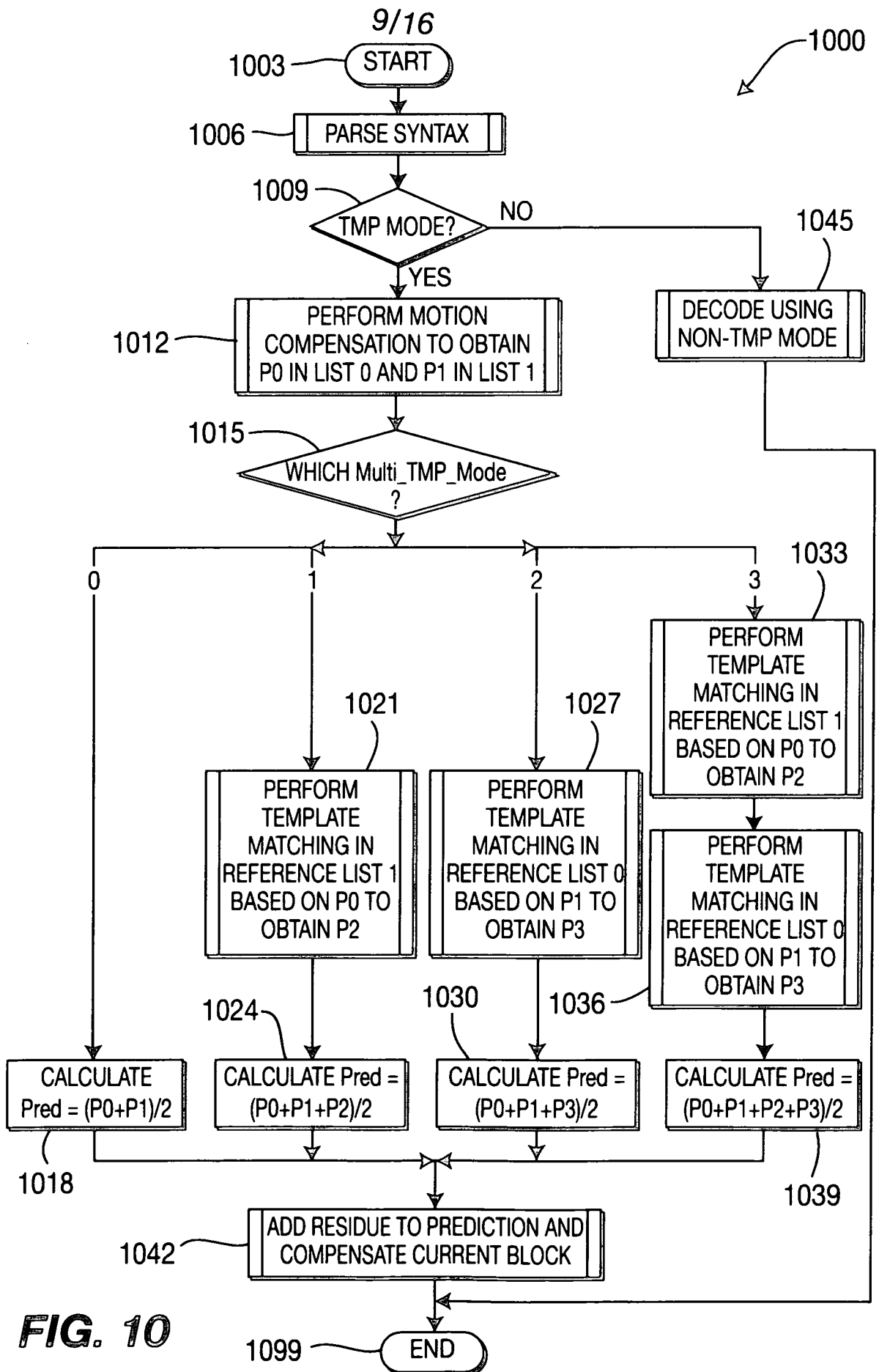


FIG. 9



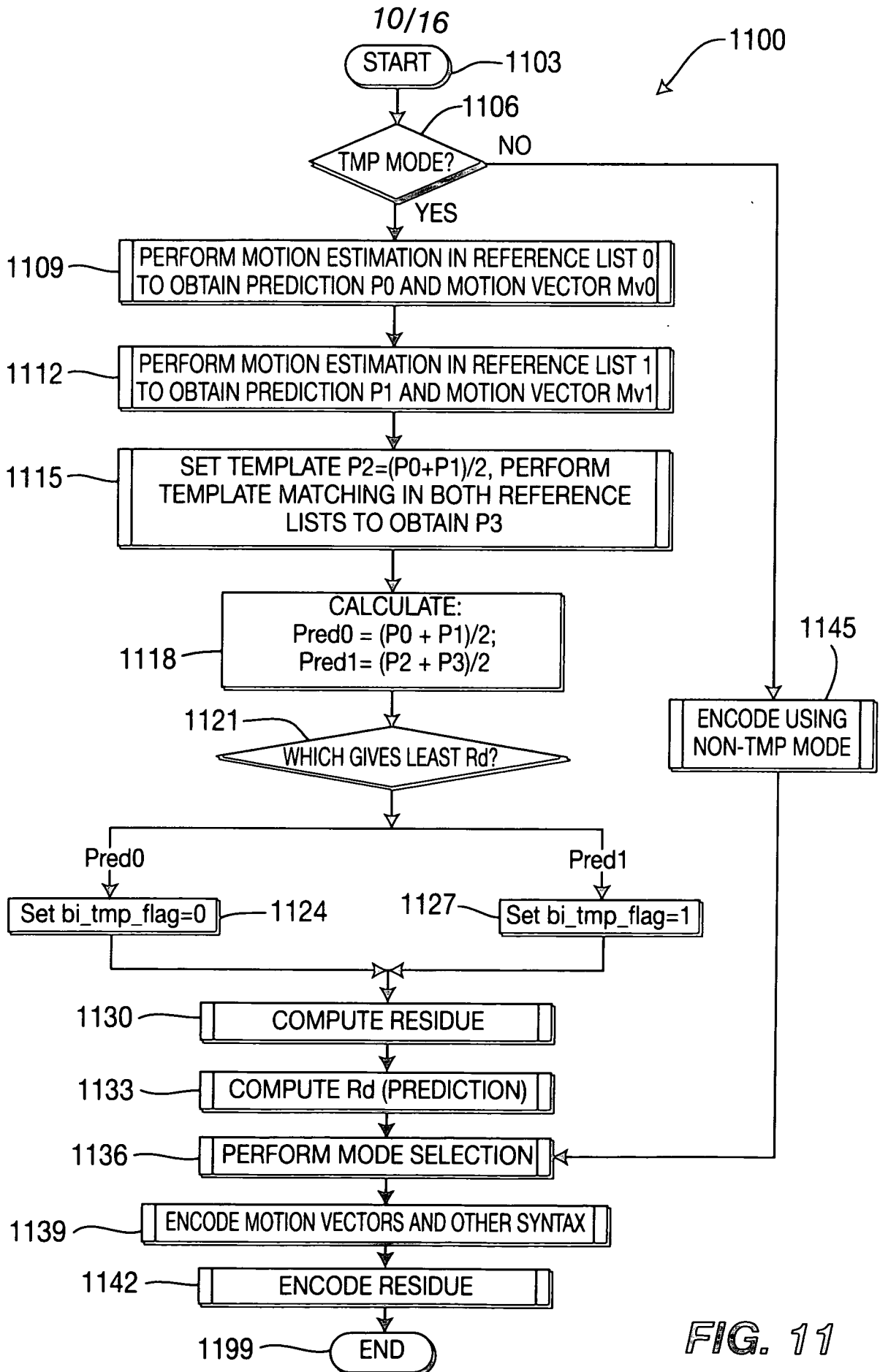


FIG. 11

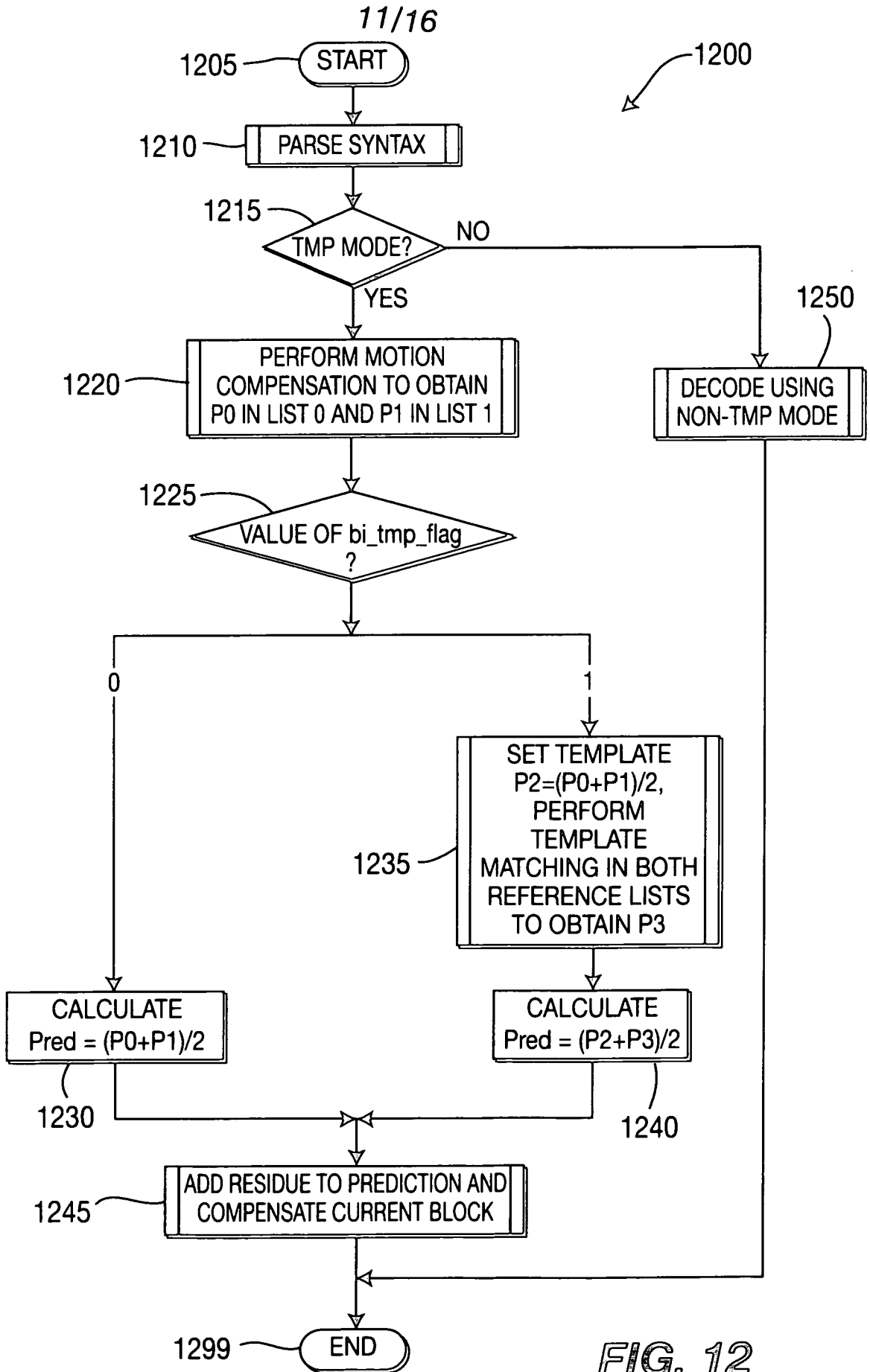


FIG. 12

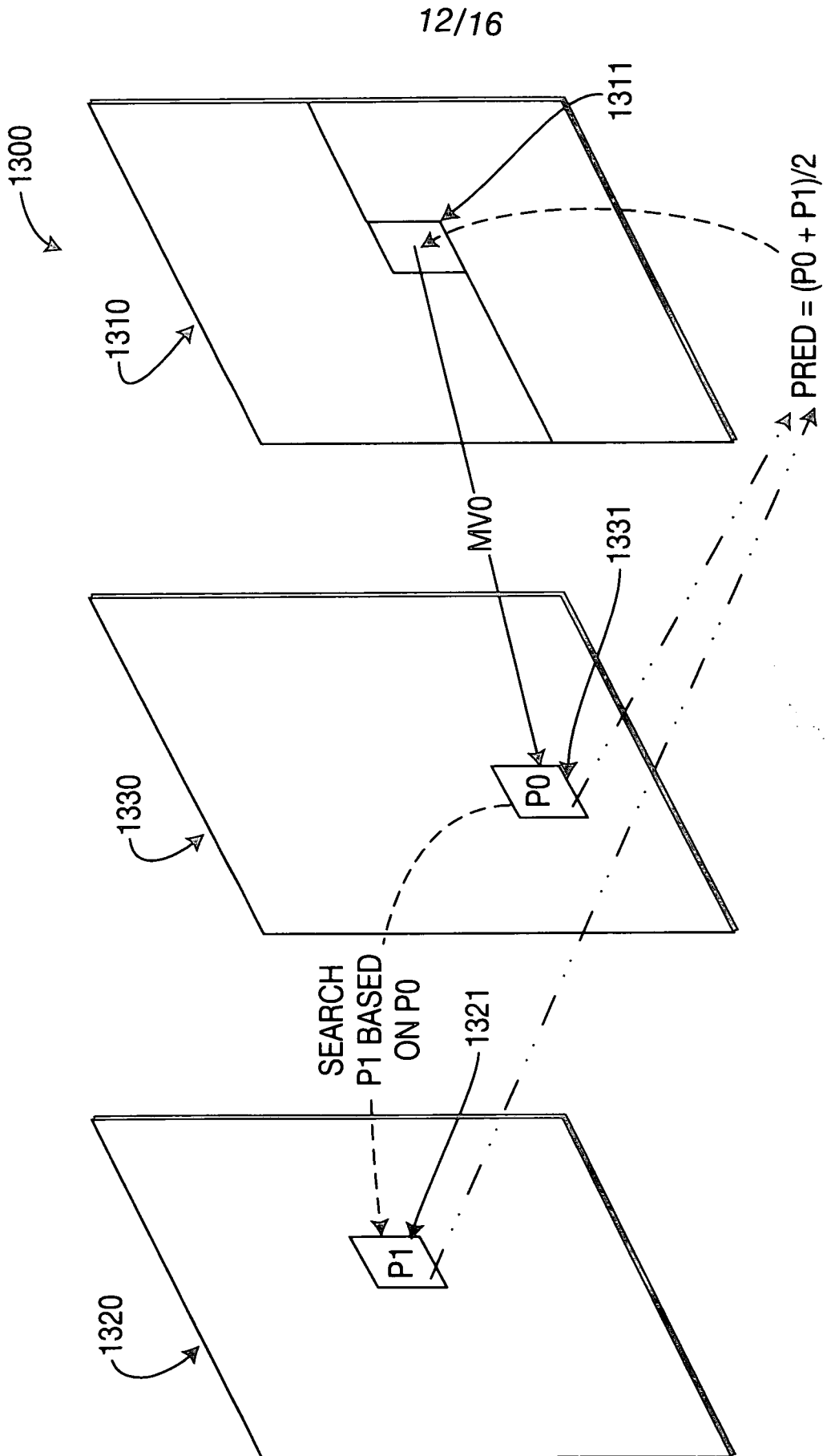


FIG. 13

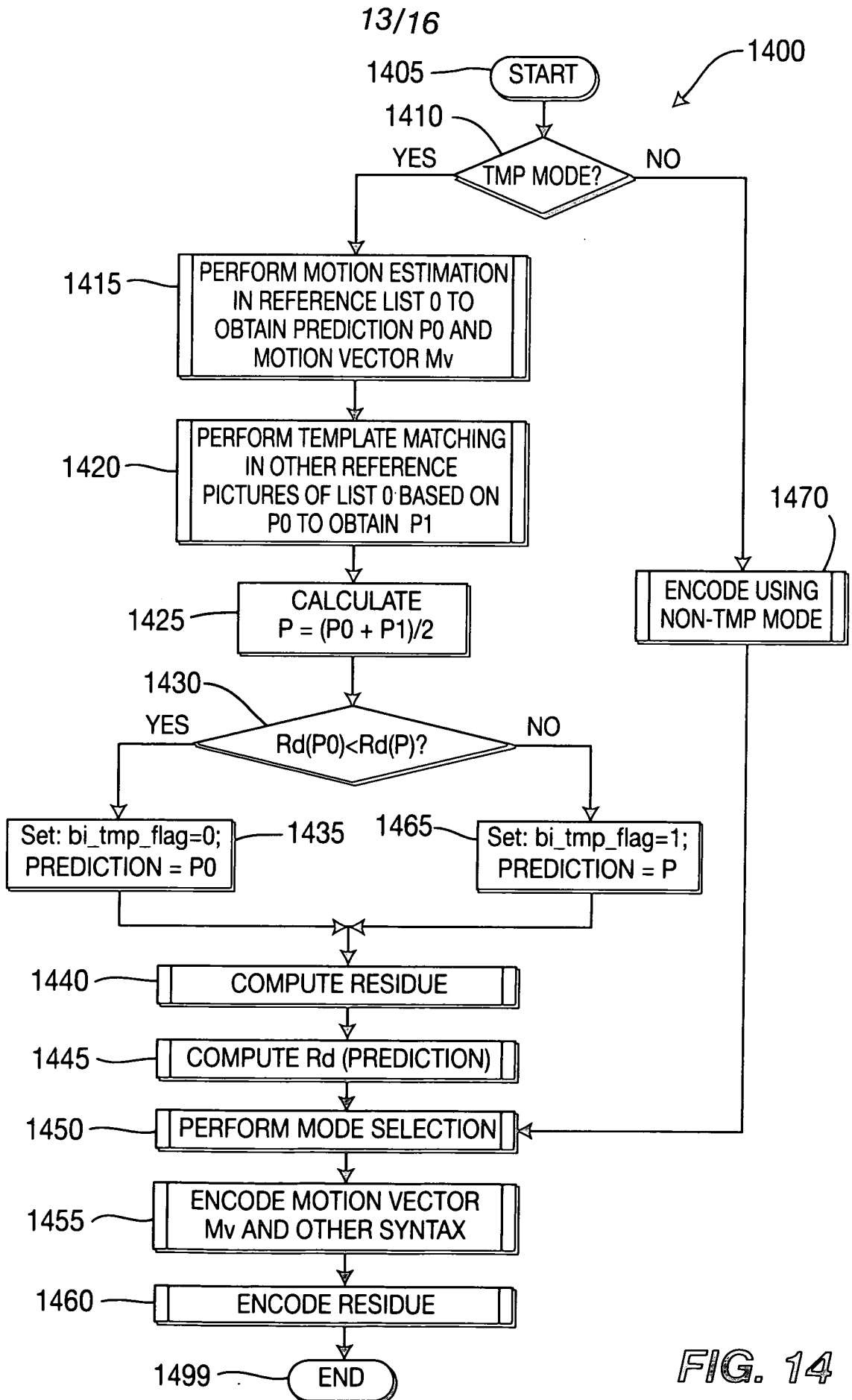


FIG. 14

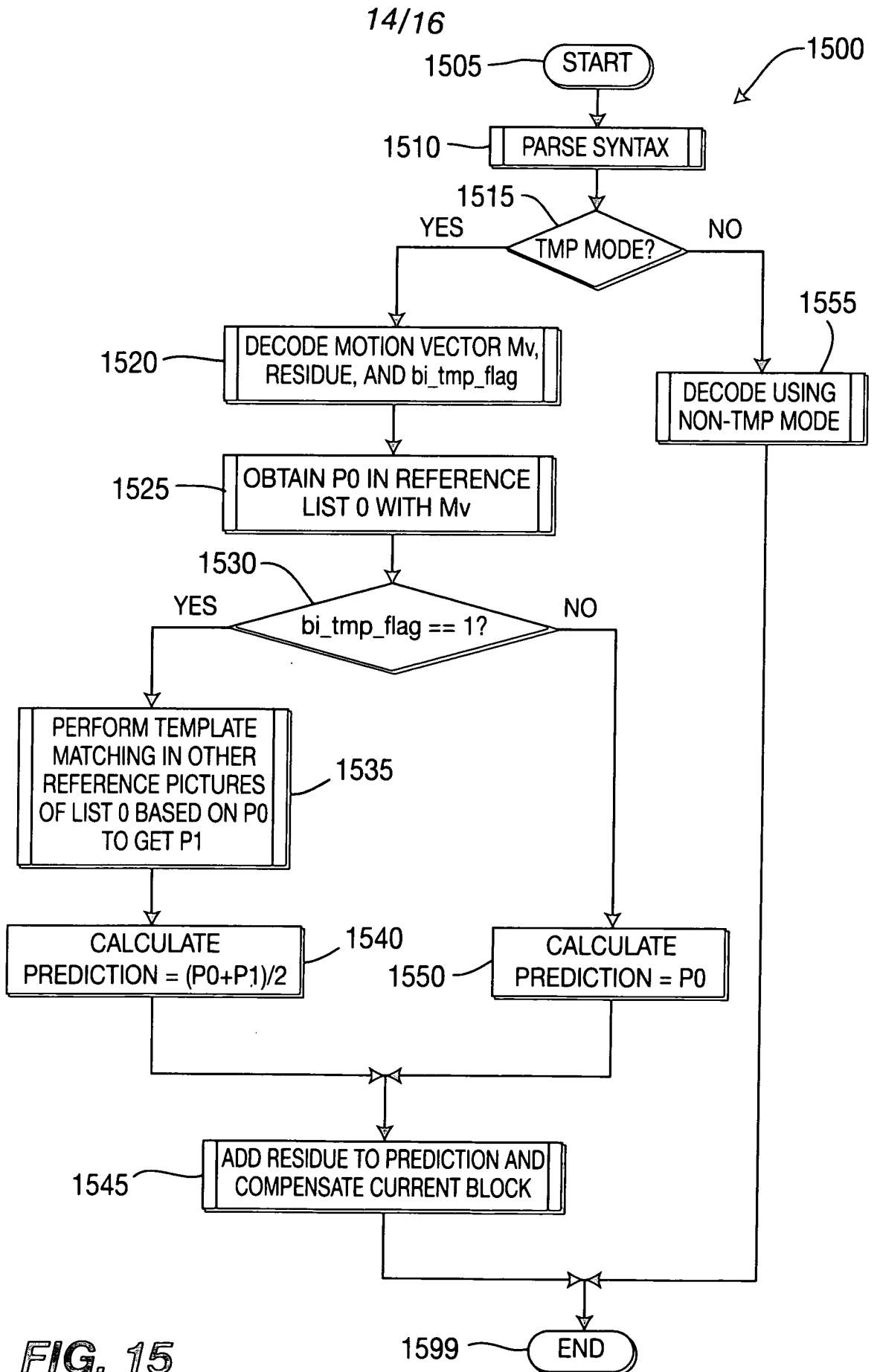


FIG. 15

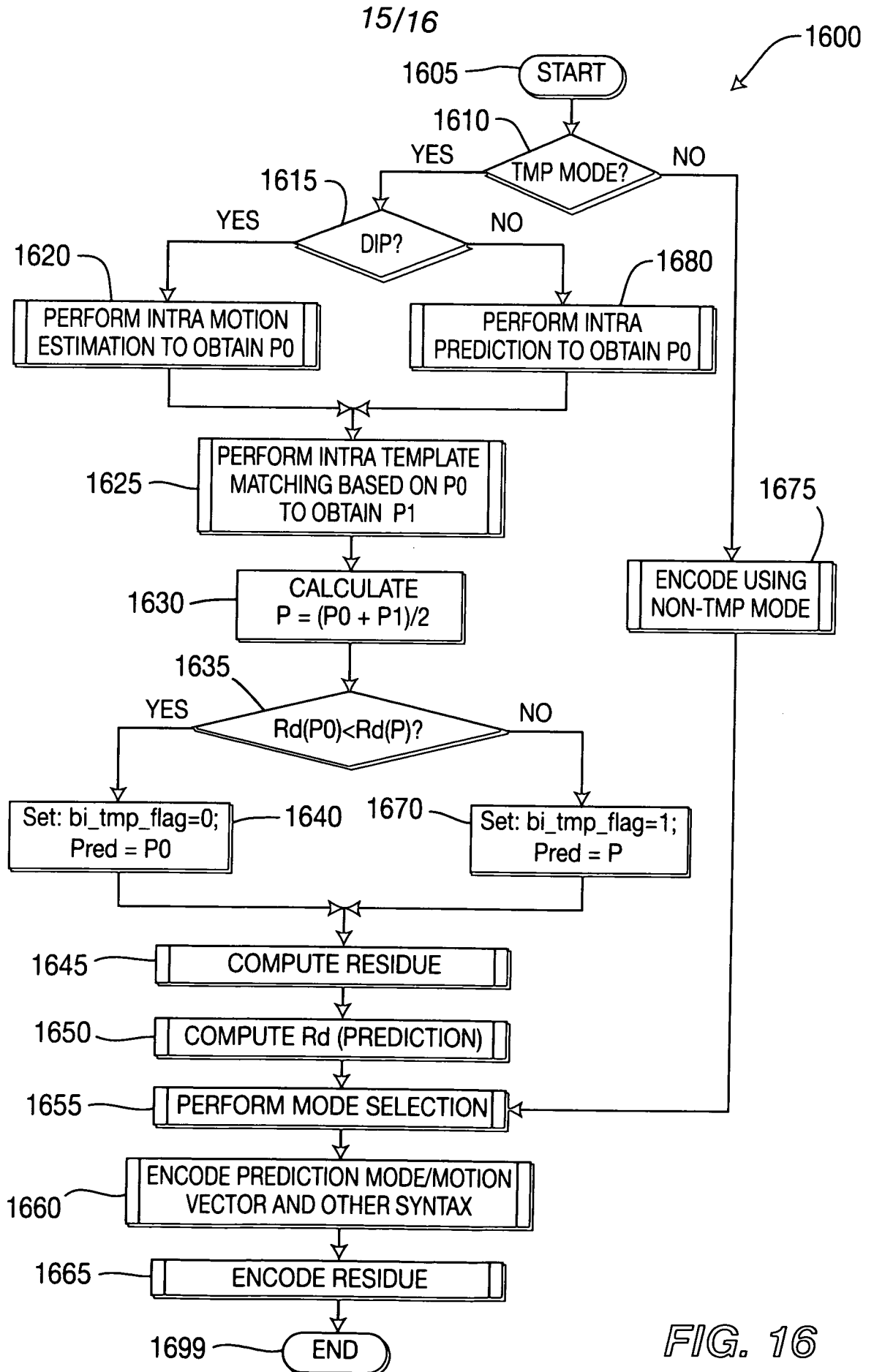


FIG. 16

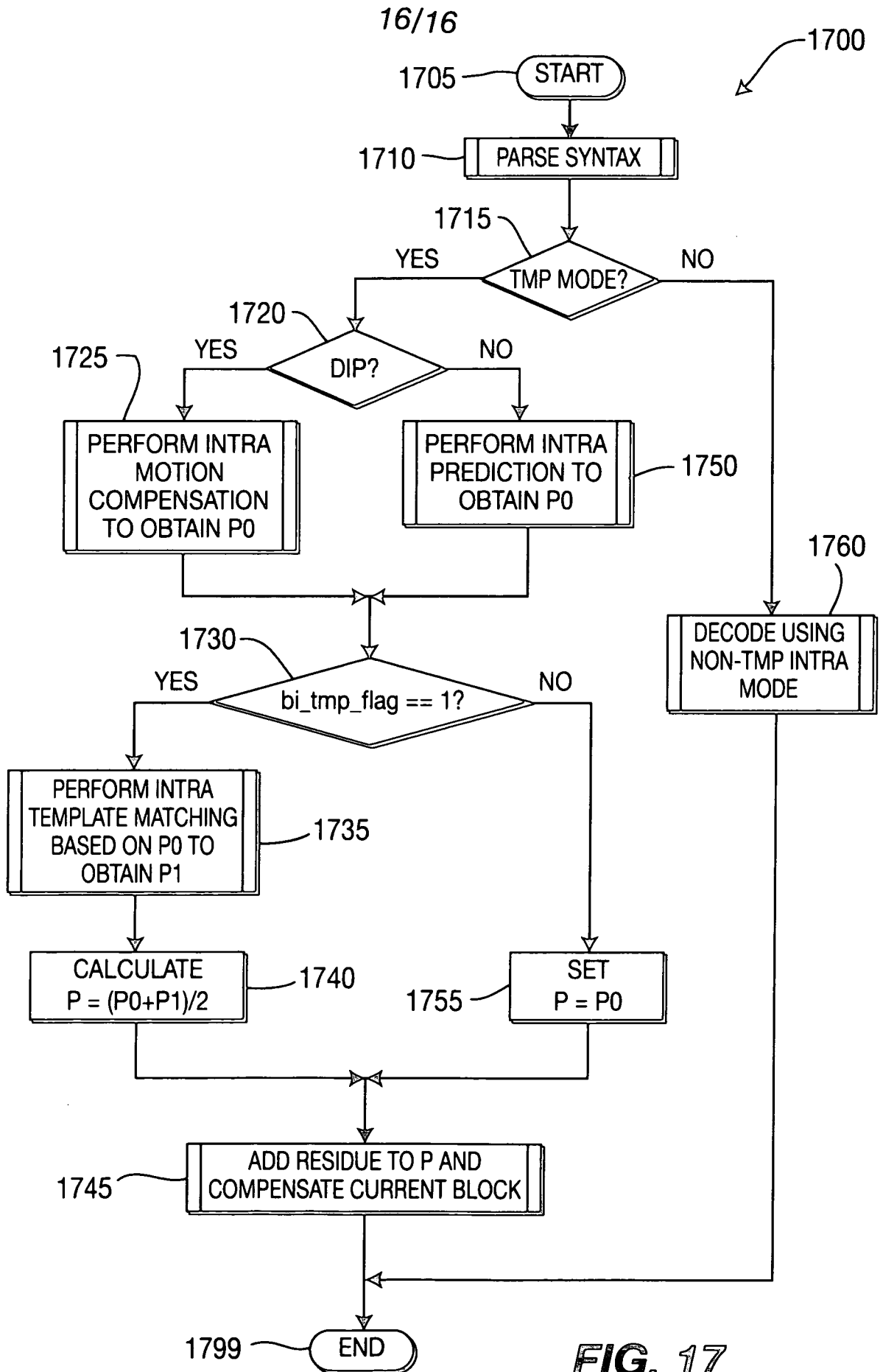


FIG. 17

INTERNATIONAL SEARCH REPORT

International application No

PCT/US2009/002177

A. CLASSIFICATION OF SUBJECT MATTER
 INV. H04N7/26 H04N7/36 H04N7/50

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, INSPEC, COMPENDEX

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	KAMP S ET AL: "Decoder-side MV derivation with multiple ref pics" 33. VCEG MEETING; 83. MPEG MEETING; 12-1-2008 - 13-1-2008; ANTALYA; (VIDEO CODING EXPERTS GROUP OF ITU-T SG.16),, no. VCEG-AH15, 12 January 2008 (2008-01-12), XP030003553	1-4, 10-14, 20-24, 30-34, 40,41
A	pages 1,2 ----- -/--	5-9, 15-19, 25-29, 35-39

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.
- * & * document member of the same patent family

Date of the actual completion of the international search

21 August 2009

Date of mailing of the International search report

16/09/2009

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

André, Thomas

INTERNATIONAL SEARCH REPORT

International application No

PCT/US2009/002177

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X A	<p>WO 99/26416 A (KONINKL PHILIPS ELECTRONICS NV [NL]; PHILIPS SVENSKA AB [SE]) 27 May 1999 (1999-05-27)</p> <p>page 2, lines 5-11 page 3, lines 20-27 page 7, line 1 - page 8, line 22</p>	<p>1-3, 10-13, 20-23, 30-33, 40, 41 4-9, 14-19, 24-29, 34-39</p>
X	<p>CHEN C-S ET AL: "Temporal Error Concealment Algorithm by Recursive Block-Matching Principle" IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, IEEE SERVICE CENTER, PISCATAWAY, NJ, US, vol. 15, no. 11, 1 November 2005 (2005-11-01), pages 1385-1393, XP011141201 ISSN: 1051-8215 pages 1386-1388, section II.A figures 2-4</p>	<p>20-24, 30-34, 40, 41</p>
A	<p>YOSHINORI SUZUKI ET AL: "Inter Frame Coding with Template Matching Averaging" IMAGE PROCESSING, 2007. ICIP 2007. IEEE INTERNATIONAL CONFERENCE ON, IEEE, PI, 1 September 2007 (2007-09-01), pages III-409, XP031158091 ISBN: 978-1-4244-1436-9 pages 409-411, sections 1-3.2</p>	<p>1-41</p>
A	<p>TANG CHENG-YUAN ET AL: "Robust two-stage approach for image motion estimation" ELECTRONICS LETTERS, IEE STEVENAGE, GB, vol. 34, no. 11, 28 May 1998 (1998-05-28), pages 1091-1093, XP006009838 ISSN: 0013-5194 abstract page 1091, first two paragraphs</p>	<p>1-41</p>
	-/--	

INTERNATIONAL SEARCH REPORT

International application No

PCT/US2009/002177

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	<p>SEILER J ET AL: "Spatio-temporal prediction in video coding by spatially refined motion compensation" PROCEEDINGS OF IEEE INTERNATIONAL CONFERENCE ON IMAGE PROCESSING, 12 October 2008 (2008-10-12), - 12 October 2008 (2008-10-12) pages 2788-2791, XP002542328 San Diego, USA abstract pages 2788-2790, sections 1-2 page 2971, section 4 figures 1-3</p> <p style="text-align: center;">-----</p>	<p>1-3, 10-13, 20-23, 30-33, 40, 41</p>

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/US2009/002177

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
WO 9926416	A	27-05-1999	DE 69834901 T2	01-02-2007
			JP 2001508632 T	26-06-2001
			US 2002025077 A1	28-02-2002
