International Journal of Academic Research and Development
ISSN: 2455-4197, Impact Factor: RJIF 5.22
www.academicsjournal.com
Volume 1; Issue 7; July 2016; Page No. 51-62

Fast motion estimation by early search termination

Shaifali Madan Arora

Assistant Professor, Electronics & Communications Engineering Department, Maharaja Surajmal Institute of Technology Janakpuri, New Delhi, India

Abstract

In this paper, a refined dynamic threshold estimation technique for zero motion prejudgment (ZMP) is proposed. In most video sequences, especially containing slow motion, large number of stationary blocks are present. Early and accurate determination of these stationary blocks reduces the computational burden without compromising the video quality in any motion estimation (ME) algorithm. Stationary blocks can be determined by comparing the block distortion with a predetermined threshold which varies depending upon the motion activity of a block. Very large or small value of this threshold results in false determination of stationary or moving blocks. So there is a need of a threshold estimation technique that follows the distortion pattern of stationary blocks in a frame. A two level dynamic threshold estimation technique is proposed in this manuscript which is based upon the distortion measure of previous stationary blocks. The performance of the proposed technique is checked by implementing ZMP before ME process in ARPS algorithm and is able to save computations by 70-95 % for slow motion sequences with very small degradation in PSNR and SSIM. Further in comparison with other fixed and dynamic threshold estimation techniques, the proposed ZMP estimation improves the PSNR and reduces the decision errors.

Keywords: Video compression, Motion Estimation, Zero Motion Prejudgment, Search Window, Spatial Correlation, Temporal Correlation

1. Introduction

Motion estimation is the most important and time consuming part of a video encoder. Pixel based, region based and block based techniques can be used to estimate the motion between the successive frames in a video. Due to simplicity and ease in hardware implementation block based motion estimation techniques are most widely used in most of the exiting video standards like MPEG-X and H.264x. Full search motion estimation [1] is the most straight forward and global optimal algorithm among the various block matching algorithms (BMAs). However, its high computational load prohibits its use in real time applications. To deal with this problem, a large number of fast BMAs have been proposed in the literature, such as three step search (TSS), new three step search (NTSS),four step search (FSS), block based gradient descent search (BBGDS), diamond search (DS) and cross diamond search (CDS) [2-7]. Most of these fast BMAs are based on the assumption that the block distortion is monotonic over the search range which means block distortion reduces monotonically as the search position approaches the global minimum point. Therefore, minimum distortion point (MDP) can be approached by following the distortion pattern without looking for all the search points in the search window. These fast BMAs follow different search patterns to speed up the search process. Further speeding of the search process is established by sub sampling the current block pixels for computing the distortion and by sub sampling the search points in the search window [8].

Another category of BMAs explore the correlation between the current block and its neighboring blocks in the spatial and temporal domains to speed the search. In these algorithms motion vector (MV) for the current block is predicted by finding the mean/median of the MVs of the neighboring blocks from spatial/temporal frames [8-10]. This predicted MV is used as a new search center and any suitable fast BMA can be applied to find the MDP. The reduction in calculations in these algorithms depends upon the accuracy of MV prediction. These algorithms require extra memory for storing MVs of neighboring blocks. To reduce memory requirement and the computational burden of prediction, authors in [11] suggested adaptive rood pattern based search (ARPS) algorithm that used the MV information of only left most block of current block for finding search center and search pattern size. Inclusion of temporal left [12] and temporal right [13] blocks are further suggested to improve the prediction of search center and search pattern size.

In applications like video telephony, video conferencing etc. there is little motion between successive frames which indicates the presence of large number of zero motion blocks in a frame [11]. The detection of these stationary blocks before starting the MDP search would lead to large savings in computations in an encoder. The decision for detecting a block to be stationary or not depends on the distortion measure of the candidate block with the collocated block in reference frame. If the calculated distortion is below a predetermined threshold than that block is categorized as stationary block and search process is terminated thereafter. The required threshold may change from block to block and frame to frame. Therefore accuracy in determination of this threshold is the key for achieving high compression without much degradation in video quality. A refined dynamic threshold prediction technique is proposed in this manuscript to decide whether a block is stationary or not. Distortion measure of the previous zero motion blocks is used for dynamically determining the threshold for the current block. The paper is organized as follows. An overview of zero motion prejudgment is given in section II. The proposed technique and its use along with ARPS algorithm are discussed in section III & IV. The proposed technique is compared with other suitable techniques for zero motion prejudgment (ZMP) in section V in terms of PSNR, average search points required per block, bits per pixel to represent residual frame, structural similarity index measurement and decision errors. Experimental results show that the proposed technique yields a very promising performance over the existing fixed and dynamic threshold prediction techniques. Concluding remarks are given in section VI.

2. Review of Zero Block Prejudgment

A. Overview

Motion estimator is the most computationally expensive part of a video encoder as it consumes 40-80% time of the encoder. A frame is divided in to blocks and motion for each block is estimated w.r.t. reference frame and is expressed in terms of MVs. For estimating the best matching block sum of absolute difference (SAD) is calculated between the current block C_{ij} and reference block R_{ij} in a predetermined search window of size W×W

$$SAD = \sum_{i=1}^{M} \sum_{j=1}^{N} |(Cij - Rij)| , \quad M \times N \text{ represents the block size} \tag{1}$$

The block which produces minimum SAD in the search window is declared as the best matching block. In full search algorithm all possible search locations are explored. Fast search algorithms reduce these search locations by following different search patterns. The computational burden in ME can be reduced by working on the reduction of search points. One way to do this is by detecting the presence of the stationary or zero motion blocks at an early stage before starting the search for best matching block in the search window. The detailed ZMP and proposed technique to do so is discussed in following subsections.

B. Zero Motion Prejudgment

In many video applications like video telephony, video conferencing etc. small motion is present between successive frames. Therefore large numbers of blocks in the current frame are same as that in the reference frame or are stationary blocks. It was found by implementing full search algorithm (block size 16×16 and search window of size 7×7) on various test video sequences containing fast, medium or slow motion activity that more than 70% blocks are stationary blocks for low motion activity video sequences like Akiyo, Salesman etc. Approximately 30% blocks are found to be stationary in medium motion activity sequences. Table 1 indicates the percentage of stationary blocks contained in various slow, medium & fast motion activity test video sequences.

Table1: Percentage of stationary blocks present in various Test sequences

Video Sequence	%age of Stationary Blocks
Akiyo	98
Salesman	97
Clair	95
News	88
Miss America	81
Hall	80
Mother Daughter	79
Stefan	38
Susie	32
Foreman	30
Highway	29
Football	26
Cactus Comb	20

The results indicate the potential of using ZMP in saving computations in MV estimation process in a video encoder. Further it was investigated in [11] that the block distortion that is calculated by finding SAD between the current and reference blocks is very low for stationary blocks as compared to moving blocks. Therefore prejudgment can be done by first finding the block distortion and comparing it with a predetermined threshold. If it is below a predetermined threshold (Ts) than that block is declared as stationary block assigning a MV (0,0) and search process is terminated thereafter, saving large amount of computations. The success and accuracy of this prejudgment depends upon the accuracy in prediction of Ts. High Ts value will result in incorrect determination of stationary blocks and hence deteriorated video quality. On the other hand low Ts will not accomplish the objective.

Authors in [11] found that the block distortion for stationary blocks vary in the range 300 to 1300 and used a fixed Ts value of 512 to detect stationary blocks and used for ZMP determination in [14] as well. However block distortion may vary from block to block and frame to frame over a very wide range. A comparison of distortion pattern of zero motion blocks in five sample test sequences is shown in fig 1.

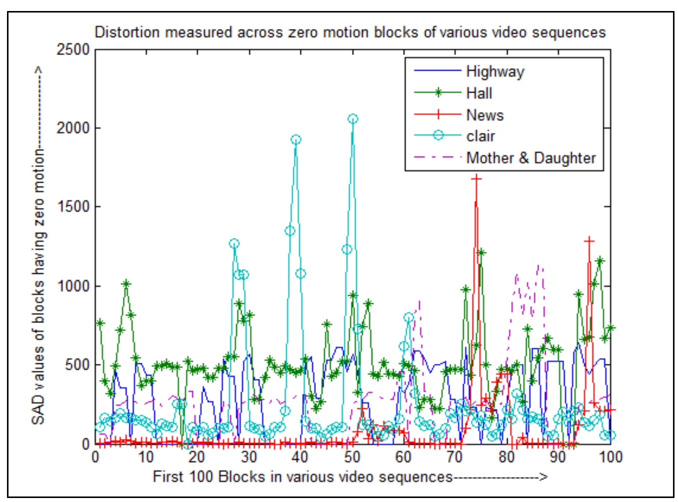

Fig 1: Distortion (SAD) Measured for stationary blocks in different video sequences

It can be observed from the figure that block distortion remains almost same for neighboring blocks as these may be part of same stationary segment of a frame but varies abruptly for different segments of the frame. In a sequence like Clair maximum distortion for stationary blocks is observed as 2000 whereas it is less than 200 for most of the blocks. Further in sequence like hall, block distortion for most of stationary blocks is more than 512 but it is very small for sequences like news. Also in Mother-Daughter sequence a segment of stationary blocks (between block no. 80 to 90) have large distortions than other stationary segments in the frame. Therefore instead of using a fixed threshold, dynamic threshold will enhance the probability of accurately finding zero motion blocks. Authors in [16] suggested a dynamic threshold estimation model as Ts = min (max (SADavg, 512), SADc)*0.75+128. In this model max operator is used to follow the changes in SAD from block to block and min operator is used to avoid the consideration of SADs of moving blocks in threshold determination model. This model is able to detect the stationary blocks which follow almost same distortion pattern. But fails to detect stationary blocks in situations as described in fig 1 for Mother-Daughter sequence between blocks 80-90 and for Hall sequence where SAD of stationary blocks varies over large intervals. Let us take one example where the average SAD of previous stationary blocks be 510 and the current stationary block and subsequent stationary block have SADs as 710,690 respectively. These blocks will be incorrectly determined as moving blocks by the above model equation.

Table 2: Variation of SAD and corresponding MVs for ten consecutive frames from Mother-Daughter sequence

Block No.	18	19	20	21	22	23	24	25	26	27	28	29	30	31
SAD	318	246	240	263	257	69	245	296	253	252	226	303	288	269
MV	(0,0)	(1,-4)	(0,3)	(0,1)	(1,-1)	(0,0)	(-1,4)	(-1,2)	(-1,4)	(-1,2)	(0,0)	(0,0)	(0,0)	(0,0)

Further table 2 illustrates the SAD and corresponding MVs for ten consecutive blocks from Mother-Daughter video sequence. From the results it is clear that if SAD is below the threshold it is not necessary to be a stationary block but has high probability to stationary. To alleviate the above drawbacks a refined dynamic threshold estimation technique is discussed in next section that enhances the accuracy in determination of stationary blocks.

3. Proposed Dynamic Threshold Estimator

The proposed technique describes the estimation of threshold Ts for ZMP. To estimate the threshold accurately, the dynamic threshold is varied according to pre-calculated parameters that follow the changes in characteristics of video sequences. Average distortion measures of previous stationary blocks are utilized for the estimation of threshold for the current block. Accuracy in determination of stationary blocks is enhanced by implementing the overall process in to two levels- Level A and Level B. Level A deals with the early detection of stationary blocks with small constant distortion whereas Level B identifies the stationary blocks with irregular and large distortions.

Level A

This level of proposed technique attempts to identify the stationary blocks that are part of same stationary segment in a frame and have almost same distortions as that of neighboring stationary blocks. The block distortion (in terms of SAD) is estimated by finding the sum of absolute difference between the current block and its collocated block in the reference frame and compared with predetermined threshold T1. If it is less than the calculated dynamic threshold then block has high probability to be stationary block from table 2. Another check is required for accurate determination of stationary block. Further it has been observed experimentally from the distortion results of various sequences that a stationary block has lowest SAD compared to its horizontal and vertical neighbors in the search window. Therefore SAD of the current block and its immediate left, right, top and bottom blocks in the search window is calculated. If the SAD of current and its collocated block is minimum, block is declared as stationary block and search is stopped thereafter otherwise another level of check at Level B is performed. The steps are summarized as

$$T1 = \max(SADa, 256) * \delta + \alpha$$

Check If $SAD_c < T1$

If SAD_c is min (SAD_c, SAD_l, SAD_r, SAD_t,SAD_b)
Than Declare the block as stationary and update SADa If abs ($T1-SAD_c$) < α Else SADa remains same,
Else Move to Level B

Where SAD_c represents the SAD between the current block and collocated reference block, SAD_l, SAD_r, SAD_t, SAD_b are the SAD between current block and left, right, top, bottom neighboring blocks respectively, T1 is the dynamic threshold for level A. Initially SADa is assigned a value of 512 from the results in [11] for fixed threshold so as to detect the first stationary block and is updated when first stationary block is encountered. Max operator helps to follow changes in SADa [16]. 256 value is taken in max operator from the results in fig 1, Sequences with very small motion have very small SAD values. Its large value increases the inaccuracy as large numbers of moving blocks which also have small SADs in slow motion sequences are determined as stationary blocks. δ (Delta) is used to control the efficiency of motion estimation process and its value is taken as 0.75 from the results of early termination model equation in [16]. Plots are drawn to show the variation of PSNR and computational complexity for different values of δ as in [16] for different activity motion sequences. As can be observed from fig 2 that with small value of δ large number of stationary blocks may remain undetected and therefore large computations are involved with only slight improvement in PSNR. Large δ would increase the prediction threshold and consequently decision errors, deteriorating the video quality by reducing the PSNR.

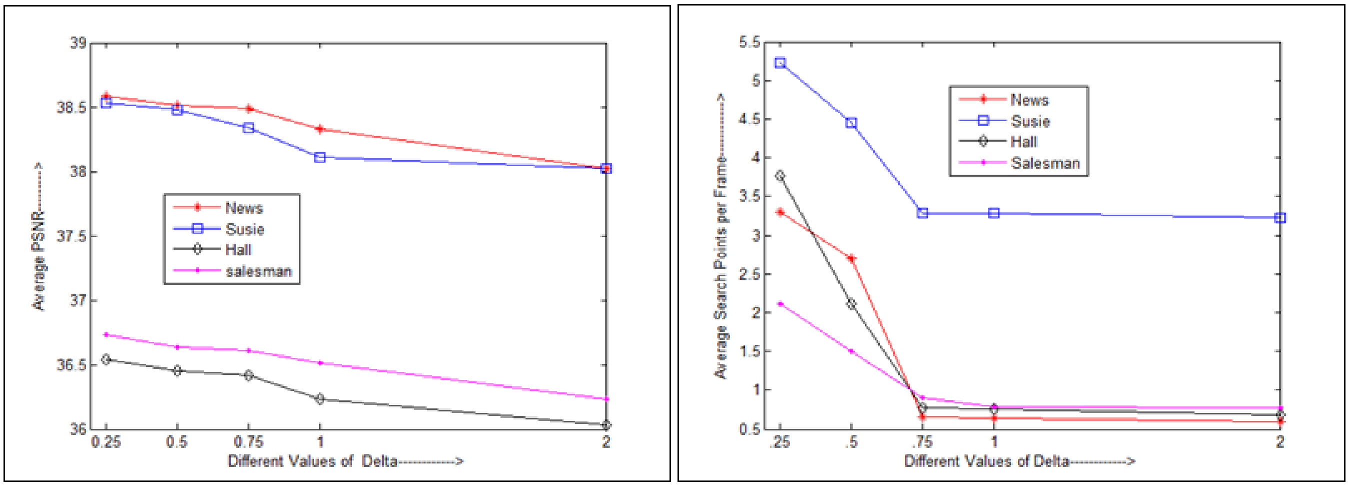

Fig 2: Effect of varying δ on average PSNR and average search points for different video

Factor α takes care of the small variations in distortions from neighboring blocks and is selected as 128 [as in 16] and also on the basis of observations of distortion pattern of various test sequences, a sample of which is shown in Fig 1.

SADa is updated by taking the average of the distortion of the previous stationary blocks. If the difference between SAD_c and T1 is more than α then that SAD_c is not considered for updating the average distortion SADa. This would save the effect of very large or very small distortion value of the current stationary block on the average variation of threshold and increase the accuracy in determination of stationary blocks. If a stationary block goes undetected at Level A, there is high probability of its detection at Level B for which another check is performed at Level B.

Level B

This level attempts to find out the stationary blocks with large distortions that are left in Level A, and compares the SAD_c with a threshold value given by T2 as

$$T2 = SAD_{0,0} * \delta + \alpha$$

Where δ and α are same as in Level A. $SAD_{0,0}$ is the SAD of the previous zero motion block. SAD_c is compared against T2 and if it is less than it, another check is performed by finding minimum from SAD_c, SAD_l, SAD_r, SAD_t, SAD_b as in Level A. If SAD_c comes out to be the minimum point declare the block as stationary block. Otherwise fast motion estimation algorithm is implemented to estimate the MVs for the block. The above procedure is summarized as

Check If $SAD_c < T2$

If SAD_c is min (SAD_c, SAD_l, SAD_r, SAD_t,SAD_b) Than Declare the block as stationary and assign $SAD_{0,0}=SAD_c$

Else

Move to Fast Motion Search Algorithm

Level B helps in detecting stationary blocks which have large or abrupt distortions.For example SAD of three consecutive stationary blocks C1, C2, C3 be 710,686,729 in certain video sequence and T1 nearly 510 and $SAD_{0,0}$ of previous stationary block be 386. Stationary block C1 will remain undetected at Level A and B. But $SAD_{0,0}$ will be updated to 710. This would help in accurate determination C2 and C3 as stationary blocks at Level B which would not be detected at Level A and in [16]. $SAD_{0,0}$ is assigned an initial value of 512.

The above stationary block determination technique can be implemented before starting the search process in any fast motion estimation algorithm. There will be slight increase in computations due to calculations involved in determining the threshold and SADs of neighboring blocks. But this small increase in computations is negligible in comparison to enhanced savings in computational burden when compared it to fixed threshold based ZMP. Further decision errors are reduced with small increase in computations in comparison to DESST (Dynamic early stop search termination) algorithm for finding ZMP in [16]. To evaluate the performance of proposed technique, adaptive rood pattern search algorithm [11] is utilized for fast motion estimation. The proposed ZMP technique is applied before the motion vector estimation process in ARPS and is discussed in next section.

4. Proposed Algorithm

Once a block is found to be moving or stationary, the next step is to estimate the MVs of the block if it is moving. The proposed dynamic zero motion detection technique is implemented before ARPS (step 5 to step 10) which is used for fast BMA, the whole procedure is summarized in terms of following steps:

1. Find the block distortion of the current block with its collocated block in the search window.
2. If it is less than the predicted threshold T1 as in Level A, Find min (SAD_c, SAD_l, SAD_r, SAD_t, SAD_b). If SAD_c comes out to be minimum distortion, declare the block as zero motion block, update the average SAD_a

as stated in level A and go to step 9 otherwise go to step 3.
3. Compare SAD_c with T2. If it is less than the predicted threshold T2 as in Level B, Find min (SAD_c, SAD_l, SAD_r, SAD_t, SAD_b). If SAD_c comes out to be minimum distortion, declare the block as zero motion block. Update SAD_c= $SAD_{0,0}$ and go to step 9 otherwise move to fast ME algorithm in step 4 onwards.
4. Find the MVs of the left adjacent block in the current frame i.e. P5 (MVx, MVy).
5. Find step size s from max magnitude of (MVx, MVy). If s is zero, search pattern is converged to a single point.
6. Create a search pattern using four horizontal and vertical points P1, P2, P3, P4 as shown in figure 3 at a distance s from center and point P5.
7. Find minimum distortion point (MDP) amongst these points, declare this MDP as new search center.
8. Create a search pattern with four horizontal & vertical points at unit distance apart from search center. Find the MDP from these five points. If MDP happen to be the center point, go to step 9. Otherwise repeat this step until MDP comes out to be the center point or search window boundary is met. Declare the MVs from the position of MDP.
9. Exit

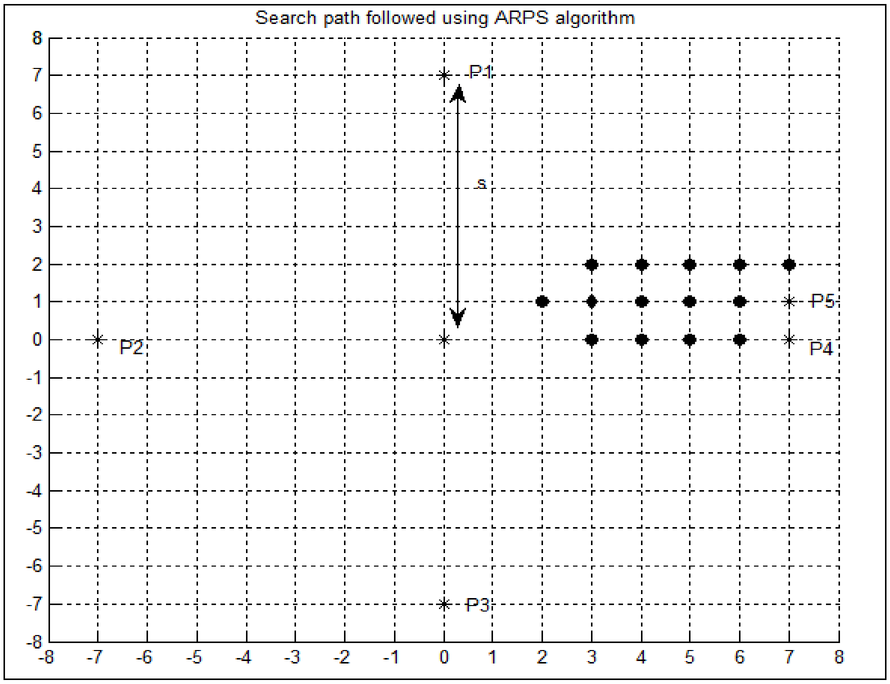

Fig 3: Search Path followed by ARPS algorithm to estimate MVs.

Flow chart of the above algorithm is shown in figure 4. Figure 3 shows motion estimation of a particular non stationary block using ARPS algorithm for a block of size 16×16. Size of the search window is taken as 7×7, which is the maximum length from MV of left adjacent block at location (1, 7). In the first iteration, five points labeled P1, P2, P3, P4, P5 and center point are processed to estimate MDP which in this case comes out at location (1, 7). It is taken as center for four points at unit size forming a search pattern as directed in step 8. These points are again explored to find MDP which results at location (1, 6). The process is repeated until MDP is the center point itself and search is stopped thereafter. For this block it comes out to be (1, 3) and is declared as the MV for this block.

5. Simulation Results

Simulations have been performed on different test video sequences to evaluate the performance of proposed stationary block prejudgment technique. Standard search window size of ±7, block size of 16×16 and15fps have been taken. Performance of the proposed dynamic threshold estimator for zero motion prejudgment is compared with fixed threshold suggested in [11] and dynamic threshold in [16]. For all the three types of threshold estimations ARPS is used as a base fast motion estimation algorithm. Comparisons are done on the basis of average number of search points per block over the entire frame, average peak signal to noise ratio, average number of bits required per pixel, to represent residual frame, structural similarity index measurement (SSIM) and decision errors. All simulations have been performed using MATLAB Version 7.6.0.324 (R2008a) on Windows 7 platform over a PC with Intel® Core™i5-2450MCPC with installed memory of 6GB and 64 bit operating system and processor speed of 2.5 GHz.

Table 3 shows the comparison of the performance of the above thresholding techniques in terms of peak signal to noise ratio (PSNR in dB) per frame which is defined as

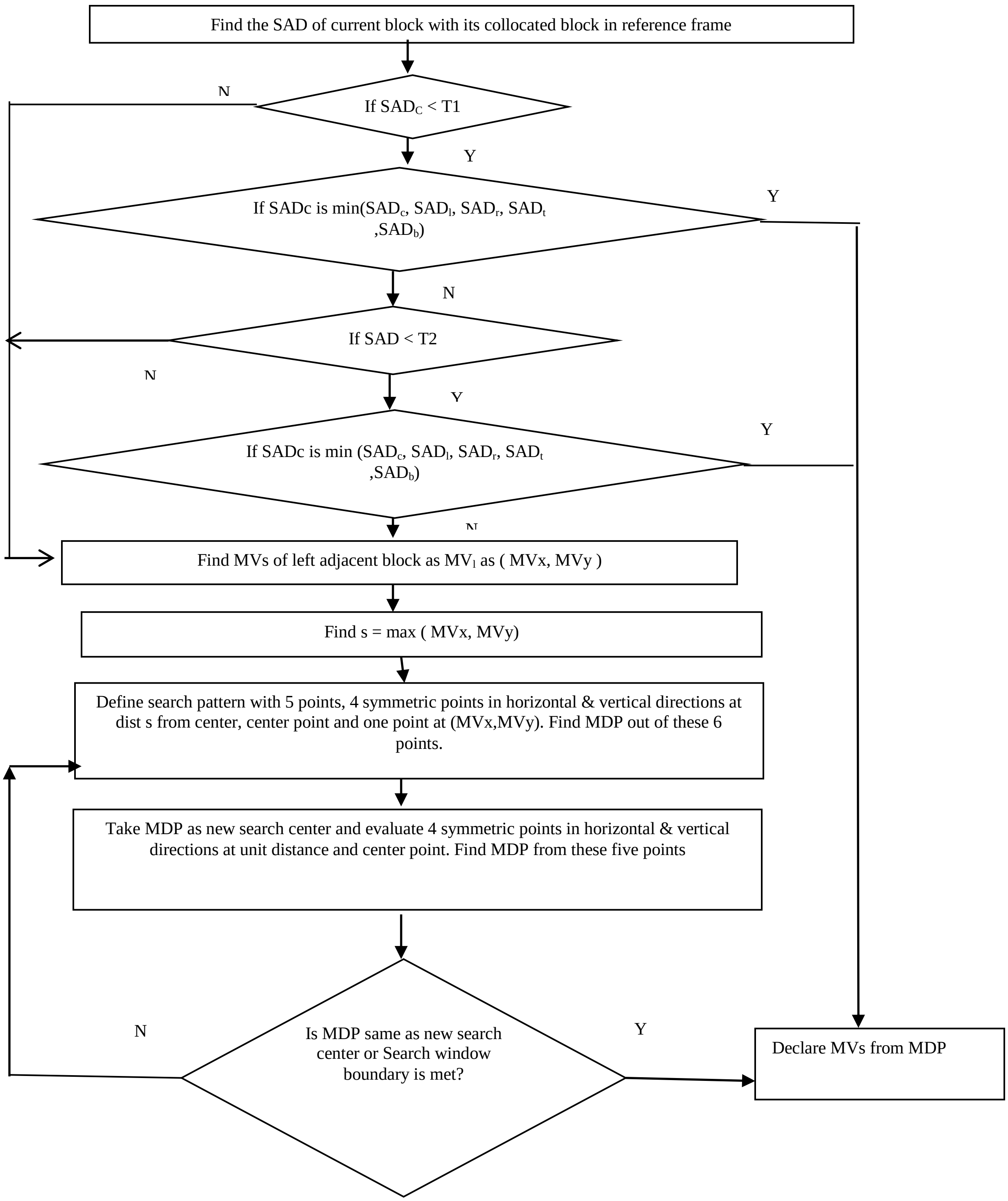

Fig 4: Flowchart of proposed Dynamic ZMP with ARPS (used as Fast BMA)

$$PSNR = 10\log_{10}\frac{Max^2}{MSE} \quad (2)$$

Where Max is the maximum possible pixel value in the video frame. All the video sequences are converted to 8 bit grayscale image frames for simulations. Therefore Max value will be 255. MSE represents the mean square error between the original frame and the generated motion compensated frame. PSNR is calculated for different types of video sequences as shown in table. FS algorithm gives the best possible matching results and therefore PSNR as it compares all the blocks in the search window with the candidate block for finding the best matching block and is therefore taken as optimum value for comparison [17] with other fast algorithms. It can be clearly observed from Table 3 that there is small degradation in video quality by estimating the zero motion blocks early. This degradation is because of the small number of moving blocks determined as stationary blocks. Accuracy in determination of stationary blocks would produce almost same video quality as that of the fast motion estimation algorithm without ZMP (ARPS algorithm is used in this manuscript).As can be observed from the table 3 that the proposed ZMP technique is able to produce better PSNR compared to fixed threshold used for ZMP or dynamic threshold in DESSTA. Average PSNR degradation calculated by performing simulation on various sequences is 0.2899dB in comparison to FS, 0.0049 dB in comparison to ARPS algorithm. It is able to produce better average PSNR by 0.0131dB w.r.t fixed thresholding technique and by 0.0246dB compared to DESSTA in [16].

Figures 5 & 6 shows the variation of PSNR over first100 of "Mother-Daughter" (slow motion sequence) and 90 frames of "Stefan" (fast motion sequence) respectively. It can be observed from above figures that proposed technique generates better video quality as compared to ZMP in [16] and very small degradation compared to FS.

Table 3: Average PSNR per frame for different video sequences

Sequence	Full Search	ARPS without ZMP	ARPS Algo with ZMP using		
			Fixed Th	DESST	PDT
Mdaughter (144X176)	41.4246	41.3494	41.3028	41.3028	41.3486
Clair (144X176)	43.3283	43.3265	43.3230	43.3230	43.3265
News (288X352)	37.7431	37.6314	37.6150	37.6150	37.6301
Foreman (144X176)	28.8384	28.6847	28.6830	28.6830	28.6847
Susie (240X352)	37.9636	37.7654	37.6702	37.6696	37.7647
Cactus comb (240X320)	20.1761	19.3306	19.3301	19.3301	19.3306
Mobile Calander (240X320)	27.9324	27.4723	27.4505	27.4164	27.4312
Hall (288X352)	36.5534	36.4514	36.4448	36.4445	36.4514
Highway(288X352)	35.6770	34.6707	34.6649	34.6649	34.6707
Stefan (288X352)	23.5804	23.1731	23.1519	23.1519	23.1609
America (144X176)	41.6898	41.6830	41.6599	41.6599	41.6829
Football (288X352)	22.5633	22.2664	22.2663	22.2663	22.2664
Salesman (144X176)	39.0926	39.0534	39.0527	39.0527	39.0530
Amelia (720X1280)	28.9998	28.7058	28.6891	28.6680	28.6973
River (720X1280)	29.9787	29.6996	29.6921	29.5761	29.6936
Average	33.0361	32.7509	32.7331	32.7216	32.7462

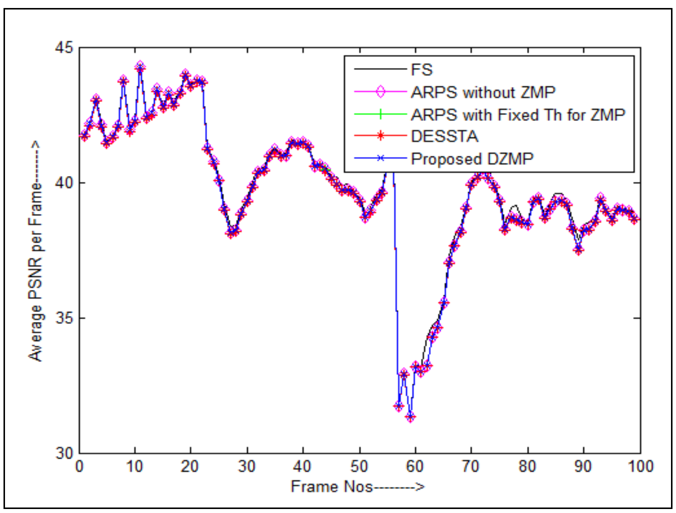

Fig 5: Average Variation of PSNR over first100 frames of "Mother-Daughter" video seq

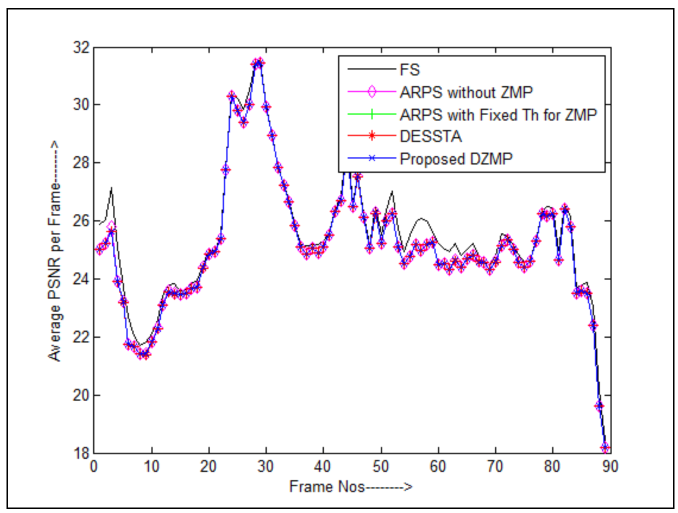

Fig 6: Average Variation of PSNR over first 90 frames of "Stefan" video seq

Table 4 shows the comparison of the proposed technique in terms of average search points required per block to find motion vectors with the FS algorithm, ARPS algorithm without ZMP and ARPS with fixed threshold for ZMP and ARPS with dynamic threshold for ZMP. Also percentage savings in computations are shown in table 4 using various thresholding techniques. It can be clearly observed from the table 4 that in ARPS algorithm the computations are reduced by 70-95% for slow motion sequences by using ZMP. As can be seen from the table that DESSTA results in better savings in computations (requires less search points) compared to proposed technique but this is at the cost of deteriorated video quality in terms of PSNR, SSIM, also more bits per pixel and decision error as can be seen from table 3,5,6 and 7. This better performance is because of the enhanced accuracy in detection of zero motion blocks by the proposed technique.

Figures 7 & 8 shows the comparison of average search points for first 100 & 90 frames of "Mother-Daughter" and "Stefan" video sequences. It can be clearly observed from figures that DESSTA requires less computations compared to proposed ZMP.

Table 4: Average search point computations per block in a frame and %age saving in search point computations per block by using ZMP

Sequence	Full Search	ARPS without ZMP	ARPS Algo with ZMP using			% Gain in computation w.r.t ARPS by using ZMP		
			Fixed Th	DESST	PDT	Fixed Th	DESST	PDT
Mdaughter (144X176)	204.2828	6.3544	2.1705	1.1684	2.3412	65.8426	81.6127	63.1562
Clair (144X176)	184.5500	4.9461	1.2999	0.2973	0.4802	73.7187	93.9892	90.2913
News (288X352)	204.2828	5.3948	1.8416	0.8407	1.0499	65.8634	84.4165	80.5387
Foreman (144X176)	184.5500	5.7170	3.9173	2.9167	3.0440	31.4798	48.9820	46.7553
Susie (240X352)	202.0485	8.2173	5.2521	4.2437	6.3875	36.0848	48.3565	22.2677
Cactus comb (240X320)	202.0485	13.8234	12.1998	11.177	12.7303	11.7453	19.1400	7.9076
Mobile Calander (240X320)	202.0485	10.8048	8.5544	6.8352	7.5013	20.8278	36.7392	30.5744
Hall (288X352)	204.2828	5.7705	4.1173	3.0968	4.5318	28.6492	46.3339	21.4661
Highway (288X352)	204.2828	8.1611	6.7775	5.7647	6.9592	16.9536	29.3637	14.7272
Stefan (288X352)	204.2828	8.1135	7.0095	6.0082	6.1389	13.6070	25.9481	24.3372
America (144X176)	184.5500	5.5568	1.2841	0.9439	2.0306	76.8914	83.0136	63.4574
Football (288X352)	204.2828	11.8493	11.7755	10.7751	10.7826	0.6228	9.0655	9.0022
Salesman (144X176)	184.5500	4.9836	1.6635	0.6630	0.9039	66.6205	86.6964	81.8625
Amelia (720X1280)	217.2486	14.2713	6.1300	3.0882	4.6422	57.0467	78.3608	67.4718
River (720X1280)	217.2486	8.4870	6.3293	3.8544	5.0360	25.4236	54.5847	40.6622
Average	200.3026	8.1634	5.3548	4.1116	4.9706	34.4045	49.6338	39.1106
Mdaughter (144X176)	204.2828	6.3544	2.1705	1.1684	2.3412	65.8426	81.6127	63.1562

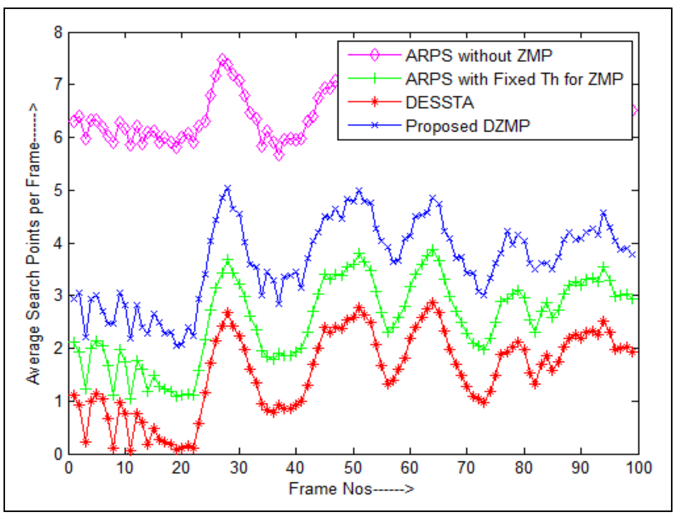

Fig 7: Average Variation of Search points over first 100 frames of "Mother-Daughter" video seq

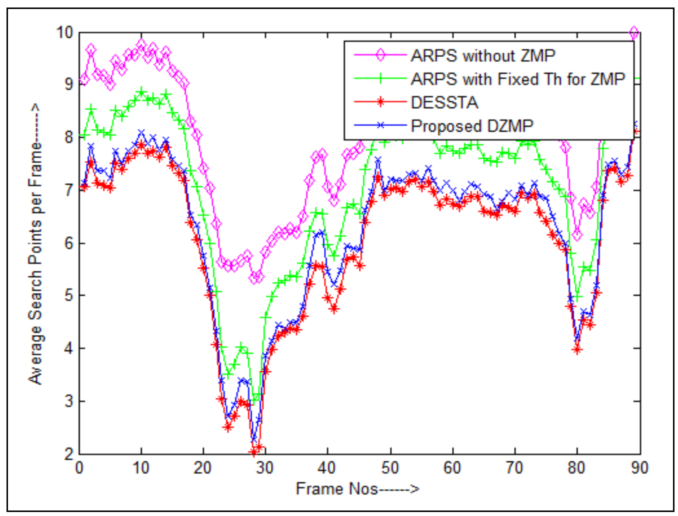

Fig 8: Average Variation of Search points over first 90 frames of "Stefan" video seq

Further table 5 shows the comparison in terms of bits per pixel [15] required to encode the residual frame. Residual frame is the difference between the actual frame and the motion compensated frame. If an algorithm is able to predict accurate motion vectors, then the difference between the actual and the motion compensated frame will be small and less number of bits will be required to encode the residual frame. From the results in table 5 it is clear that FS algorithm gives optimum MVs and therefore requires least number of bits as compared to other algorithms to represent a pixel in residual frame. The proposed technique requires 1.0078 times more bits compared to FS, almost times more bits compared to ARPS algorithm but requires less 1.0037 times less bits compared to the DESSTA in [16].

Table 5: Average bits required to represent a pixel in a residual frame

Sequence	Full Search	ARPS without ZMP	ARPS Algo with ZMP using		
			Fixed Th	DESST	PDT
Mdaughter (144X176)	3.0092	3.1208	3.0296	3.0296	3.0180
Clair (144X176)	2.1125	2.1136	2.1135	2.1126	2.1135
News (288X352)	2.4022	2.4024	2.4056	2.4057	2.4056
Foreman (144X176)	2.6686	2.6729	2.6721	2.6721	2.6729
Susie (240X352)	3.3496	3.3765	3.4264	3.4267	3.3765
Cactus comb (240X320)	5.6867	5.7855	5.7977	5.7979	5.7858
Mobile Calander (240X320)	1.8145	1.8172	1.8201	1.8255	1.8196
Hall (288X352)	3.7265	3.7308	3.7326	3.7326	3.7308
Highway (288X352)	3.8397	3.8937	3.8954	3.8954	3.8937
Stefan (288X352)	4.9795	5.0127	5.0144	5.0144	5.0134
America (144X176)	2.8767	2.8791	2.8922	2.8923	2.8792
Football (288X352)	5.9173	5.9685	5.9685	5.9685	5.9685
Salesman (144X176)	2.8598	2.8598	2.8599	2.8599	2.8598
Amelia (720X1280)	2.4700	2.5607	2.6176	2.6498	2.5613
River (720X1280)	1.9042	1.9070	1.9084	1.9088	1.9071
Average	3.3078	3.3401	3.3436	3.3461	3.3337

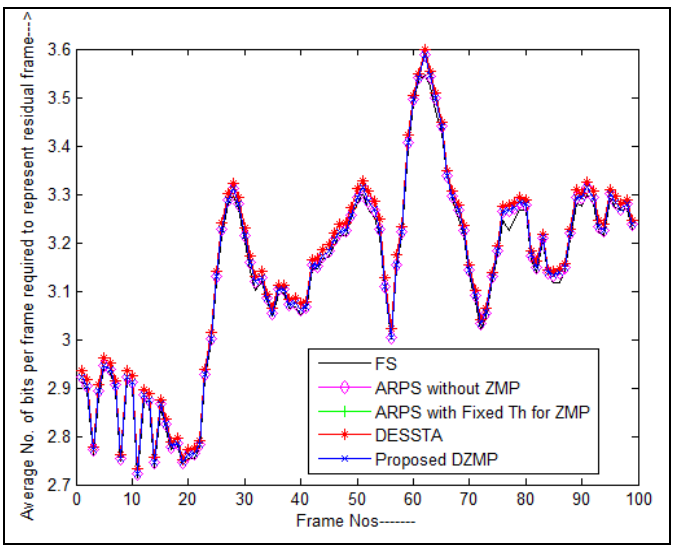

Fig 9: Variation of average bits per pixel over first 100 frames in "Mother-Daughter" video seq

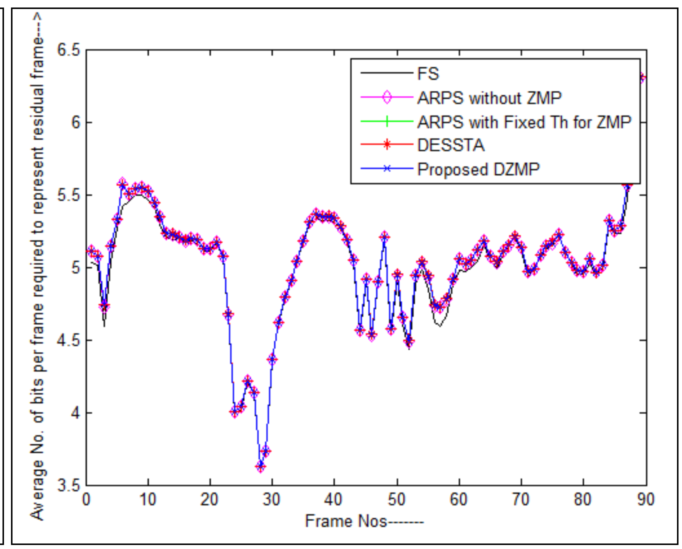

Fig 10: Variation of average bits per pixel over first 90 frames in "Stefan" video

Figures 9 & 10 show the variation of bits per pixel over first 100 for "Mother-Daughter" and 90 frames of "Stefan" video sequences respectively. As it can be observed from the figures that less number of bits per pixel are required in each frame for the proposed ZMP compared to fixed thresholding or DESSTA.

Table 6 shows the performance of above stated algorithms in terms of structural similarity index measurement (SSIM). SSI is a method for measuring the similarity between two images. SSIM considers the image degradation as perceived change in structural information and is different from MSE (Mean Square error) and PSNR as these approaches estimate perceived errors [18]. Structural information is the idea that pixels have strong interdependencies especially when they are spatially close. These dependencies carry important information about the structure of objects in a visual scene. In this manuscript, SSIM between the original frame and the motion compensated frames is measured. A video frame is divided into fixed size blocks of size 16X16 for simulations. SSIM between the two blocks, block A in the original frame and block B in the motion compensated frame, can be measured as:

$$SSIM(A,B) = \frac{2(u_A u_B + c1)(2\sigma_{AB} + c2)}{(u_A^2 + u_B^2 + c1)(\sigma_A^2 + \sigma_B^2 + c2)} \quad (3)$$

Where u_A and u_B are average of block A and block B respectively. σ_A^2 and σ_B^2 are covariance of Block A and B respectively. σ_{AB} is covariance between A and B. $c_n=(k_n l)^2$ is a constant calculated from constants $K_1=0.01$, $K_2=0.03$ and l =255(dynamic range of pixel values).

Average SSIM is measured over test video sequences and the results are shown in table 5. Simulation results show the better SSIM produced by proposed dynamic technique as compared to DESSTA in [16] by.009. Small quality deterioration by 0.004 and 0.0002 compared to FS and ARPS respectively is because of the small errors involved in ZMP. Figures 11 & 12 show the variation of SSIM over first 100 frames of "Mother-Daughter" and 90 frames of "Stefan" video sequences.

Table 6: Average Structural Similarity Index Measurement per frame

Sequence	Full Search	ARPS without ZMP	ARPS Algo with ZMP using		
			Fixed Th	DESST	PDT
Mdaughter (144X176)	0.9711	0.9706	0.9701	0.9701	0.9706
Clair (144X176)	0.9921	0.9921	0.9921	0.9921	0.9921
News (288X352)	0.9835	0.9830	0.9829	0.9829	0.9829
Foreman (144X176)	0.9094	0.9089	0.9089	0.9089	0.9089
Susie (240X352)	0.9593	0.9577	0.9565	0.9565	0.9577
Cactus comb (240X320)	0.6778	0.6496	0.6489	0.6489	0.6496
Mobile Calander (240X320)	0.7745	0.7741	0.7735	0.7725	0.7729
Hall (288X352)	0.9465	0.9465	0.9464	0.9463	0.9464
Highway (288X352)	0.9063	0.9005	0.9003	0.9003	0.9005
Stefan (288X352)	0.8581	0.8436	0.8426	0.8426	0.8432
America (144X176)	0.9806	0.9806	0.9804	0.9804	0.9806
Football (288X352)	0.7020	0.7020	0.7020	0.6948	0.7000
Salesman (144X176)	0.9716	0.9714	0.9714	0.9714	0.9714
Amelia (720X1280)	0.9308	0.9308	0.9284	0.9268	0.9288
River (720X1280)	0.9374	0.9372	0.9364	0.9326	0.9357
Average	0.9001	0.8966	0.8961	0.8951	0.8961

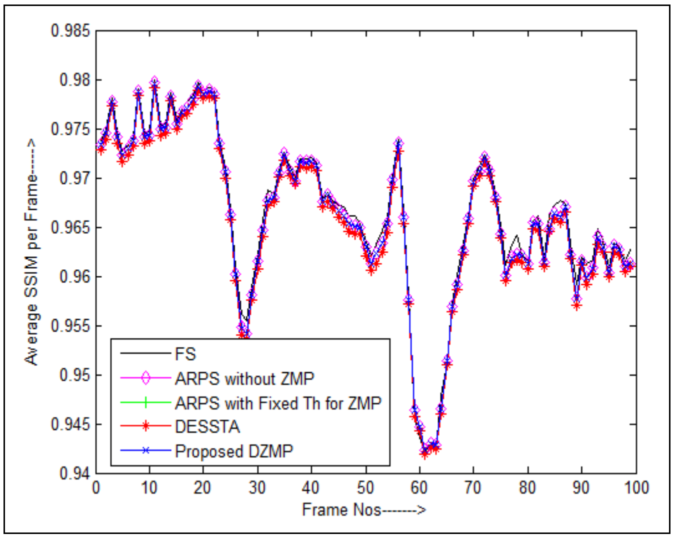

Fig 11: Average Variation of SSIM over first 100 frames of "Mother-Daughter" video sequence

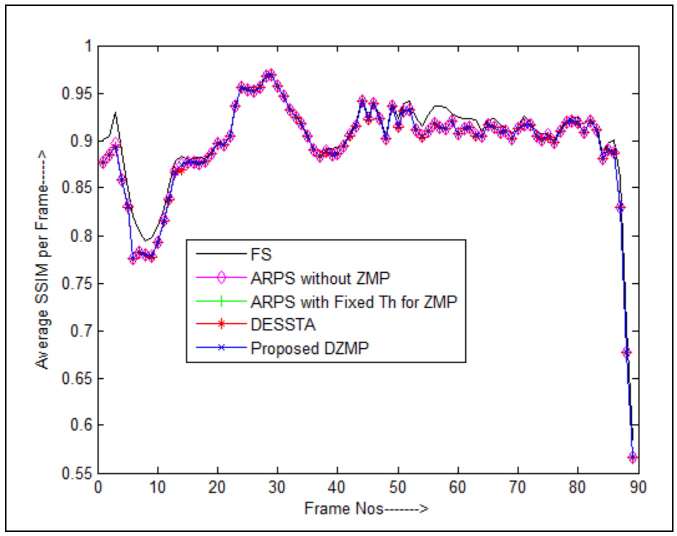

Fig 12: Average Variation of SSIM over first 90 frames of "Stefan" video seq

The accuracy in prediction of threshold value plays very important role for early detection of stationary blocks. Errors may result due to the incorrect prediction of stationary blocks as

Table 7: Average decision error per frame

Sequence	ARPS Algo with ZMP using		
	Fixed Th	DESST	PDT
Mdaughter (144X176)	0.2402	0.2402	0.0743
Clair (144X176)	0.0060	0.0060	0.0040
News (288X352)	0.0093	0.0093	0.0091
Foreman (144X176)	0.0253	0.0253	0.0253
Susie (240X352)	0.4077	0.4094	0.0900
Cactus comb (240X320)	13.9017	14.0818	2.6263
Mobile Calander (240X320)	0.7222	1.1132	0.5499
Hall (288X352)	0.0370	0.0559	0.0534
Highway (288X352)	0.5800	0.5800	0.2587
Stefan (288X352)	0.1015	0.1015	0.0692
America (144X176)	0.0837	0.0837	0.0237
Football (288X352)	0.1775	0.2271	0.1881
Salesman (144X176)	0.0011	0.0011	0.0009
Amelia (720X1280)	6.2350	5.7458	4.3470
River (720X1280)	0.4075	1.0584	0.3465
Average	1.5290	1.5826	0.5778

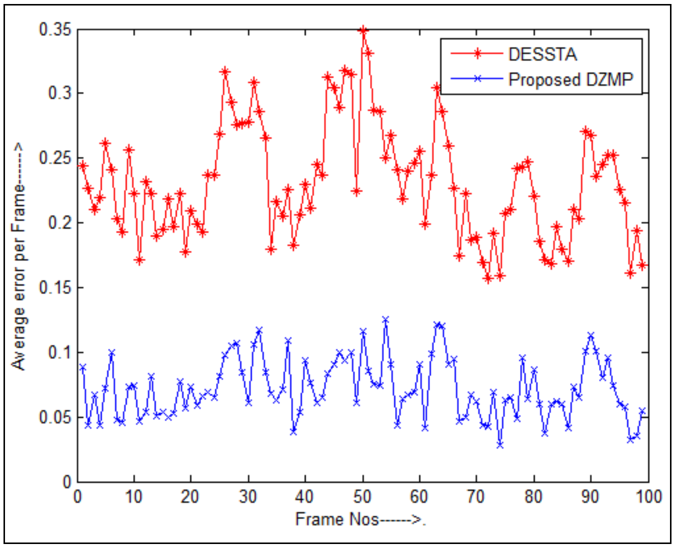

Fig 13: Variation of decision error for first 100 frames in "Mother-Daughter" video seq

moving or moving blocks as stationary i.e. known as decision errors. This is the most important parameter to judge the performance of ZMP technique and is defined by equation

$$Decision\ Error = \frac{N_{sm} + N_{ms}}{N_s} \quad (4)$$

Where N_{sm} is number of stationary blocks incorrectly determined as moving blocks, N_{ms} is total number of moving blocks incorrectly determined as stationary blocks, N_s is actual total number of stationary blocks determined using Full Search algorithm. From the table 7 it is clear that the proposed technique produces least decision error compared to fixed threshold and DESST in [16] for prediction of stationary blocks.

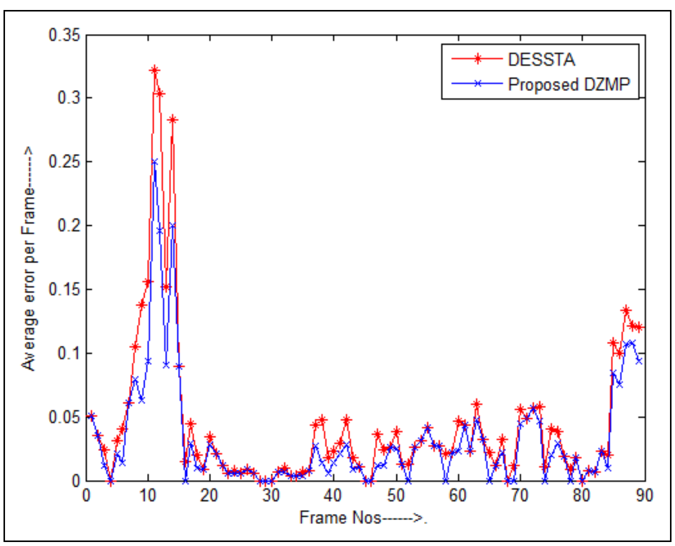

Fig 14: Variation of decision error for first 90 frames in "Stefan" video sequence

The decision error results are shown in figures 13 & 14 for first 90 frames of "Mother-Daughter" and "Stefan" video sequences. It can be clearly observed from the figures that with proposed dynamic threshold the decision errors are reduced compared to the DESS threshold technique in [16].

6. Conclusion

In this paper, a refined dynamic zero motion prejudgment technique is proposed to enhance the accuracy in determination of stationary blocks. The experimental results have shown dynamic threshold prediction is better able to estimate stationary blocks than the fixed threshold. Further the proposed technique can produce better video quality in terms of PSNR, SSIM with fewer bits per pixel to represent residual frame and reduced decision errors compared to DESST technique suggested in [16]. ARPS algorithm is the fast motion estimation algorithm used in this paper. All the ZMP techniques are applied before the motion estimation process in ARPS. However the proposed technique can be applied before any fast motion estimation algorithm for zero motion prejudgment.

7. References

1. Jain JR, Jain AK. Displacement measurement and its application in inter-frame image coding. IEEE Transactions on Communications, 1981; 29(12):1799-1808.
2. Koga T, Iinuma K, Hirano A, Iijima Y, Ishiguru T. Motion compensated inter-frame coding for video conferencing. In: Proceedings of NTC81. New Orleans, 1981; G5.3.1-G5.3.5
3. Li R, Zeng B, Liou ML. A new three step search algorithm for block motion estimation. IEEE Transactions on Circuits and Systems for Video Technology 1994; 4(4):438-442.
4. Man Po L, Ma WC. A novel four step search algorithm for fast block motion estimation. IEEE Transactions on Circuits and Systems for Video Technology 1996; 6(3):313-317.
5. Kuo L, Feig E. A block based gradient descent search algorithm for block motion estimation in video coding. IEEE Transactions on Circuits and Systems for Video Technology, 1996; 6(4):419-422.
6. Zhu S, Ma K. A new diamond search algorithm for fast block matching motion estimation. IEEE Transactions on Image Processing, 2000; 9(2):287-317
7. Cheung C, L Po. Novel cross diamond hexagonal search algorithms for fast block motion estimation. IEEE Transactions on Multimedia, 2005; 7(1):16-22.
8. Wang Y, Wang Y, Kuroda H. A globally adaptive pixel decimation algorithm for block motion estimation. IEEE Transactions on Circuits and Systems for Video Technology, 2000; 10(6):1006-1011.
9. Junavit C, Kuo J. Fast motion vector estimation using multi-resolution spatio-temporal correlations. IEEE transactions on Circuits and systems for Video Technology, 1997; 7(3):477-488.
10. Luo LJ, Zou C, Gao XQ. A new prediction search algorithm for block motion estimation in video coding. IEEE Transactions on Consumer Electronics, 1997; 43(1):56-61.

11. Nie Y, Ma KK. Adaptive rood pattern search for fast block motion estimation. IEEE Transactions on Image processing, 2002; 11(12):1442-1448.
12. Purwar RK, Rajpal N. A fast block motion estimation algorithm using dynamic pattern search. Springer journal on Signal Image and Video Processing 2013; 7(1):151-161.
13. Arora SM, Rajpal N, Purwar RK. New Dynamic Pattern Search Based Fast Motion Estimation Algorithm. Proceedings of International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2014, 979-985.
14. Luo J, Yang X, Liu L. A fast motion estimation algorithm based on adaptive pattern and search priority. Springer journal on Multimedia Tools Applications, 2014. DoI: 10.1007/s11042-014-2280-z
15. Gonzalez R, Woods R, Eddins S. Digital Image Processing using MATLAB. Pearson Education India, 2009, 2nd edition.
16. Ismail Y, McNeely B, Shaaban M, Mahmoud M, Bayoumi MA. Fast Motion Estimation System Using Dynamic Models for H.264/AVC Video Coding. IEEE Transactions on Circuits and Systems for Video Technology, 2012; 22(1):28-42.
17. VCEG, Calculation of average PSNR differences between RD-Curves, Technical Report M33, Video Coding Experts Group (VCEG), Austin, Texas, USA, 2001.
18. Chow H, Liou K ML. A universal image quality index. IEEE Signal Processing Letters 2002; 9(3):81-84.