2-D MOTION ESTIMATION OF VIDEO SEQUENCES WITH INFLUNCE OF ZERO-MEAN GAUSSIAN NOISE

S. USAMA [1)], B. ŠIMÁK [1)], AND M. KLÍMA [2)]
1) Department of Telecommunications.
2) Department of Radioelectronics.
CTU in Prague, Faculty of Electrical Engineering.
Technicka 2,166 27 Prague 6, CZECH REPUBLIC

Abstract

This paper presents the performance of four motion estimation algorithms with present of zero-mean Gaussian noise. The performance of full search (FS) algorithm, three-step search (TSS) algorithm, new three-step search (NTSS) algorithm, and four-step search (4SS) algorithm with influence of Gaussian noise has been described for three of famous video sequences. The critical signal-to-noise ratio (SNR) is introduced for each algorithm. The worst performance is often in case of full search algorithm (FS).

Keywords

Motion compensation, Full search, Three-step search, New Three-step search, Four-step search, and Peak signal-to-noise-ratio.

1. Introduction

Motion compensation is a predictive technique for exploiting the temporal redundancy between successive frames of video sequence.

There have been numerous attempts at motion compensated image coding in the past, each with both advantages and disadvantages. An early method by Giorda and Racciu [1] divides the image into a number of rectangular blocks and determines a displacement vector for each block via correlation. In the method of pel-recursive displacement estimation by Netravali and Robbins [2], single pixels of the image are operated on and the displacement for each is determined. The most effective technique for motion estimation is the block matching algorithms (BMA). The full search algorithm (FS) is the most obvious candidate for search technique for finding the best possible weight in the search area. It searches all possible locations inside the search window "W" in the reference frame to provide an optimal solution. To locate the best match by this algorithm, $(2W+1)^2$ evaluations of the matching criterion are required. Hence, the full search (FS) algorithm actually has a complexity of $O(W^2)$. To reduce the computational complexity, Koga et al [3] use a three-step motion vector direction search (TSS) to compute displacements up to 6 pel/frame. This method, for W=6 pel/frame, search 25 positions to locate the best match. Li, Zeng, and Liou [4] modified the three-step search algorithm to become efficient for the estimation of small motions. The basic idea of new three-step search algorithm (NTSS) is that it employs a center-biased checking point pattern in the first step. The NTSS algorithm requires 33 block matches as compared to 25 matches needed in TSS. Po and Ma [5] use a four-step search (4SS) algorithm with center-biased checking point pattern. This method reduces the worst-case computational requirement from 33 to 27 search positions and the average computational requirement from 21 to 19 search positions as compared with NTSS. Many fast block-matching techniques for motion vector estimation have been developed and evaluated in the literature [6-10].

Once the motion estimation has been performed for a pair of frames, the motion vectors are used to warp the pixels of the previous frame towards the current frame and hence reconstruct the current frame. Due to the noise effect, motion vectors are not sufficient for accurate estimation, and there is critical signal-to-noise ratio for good estimation.

Gaussian type noise is a common type of noise that appears within images. This kind of noise is due to the discrete nature of radiation, i.e. the fact that each imaging system is recording an image by counting photons.

The criteria of failure point for motion estimation algorithms is described in section 2. This criteria is based on the peak-signal-to-noise ratio (PSNR) function, proposed by J. R. Jain, and A. K. Jain in [11]. The performance evaluations for each algorithm are reported in section 3. The noise variance is changed to study the effect of variance on PSNR and the results are reported in section 4. The effect of changing block size from 8 to 16 is discussed in section 5. Conclusions are presented in section 6.

2. Criterion of Failure Point for Motion Estimation Algorithms

The coding performance is measured in decibels (dB) using the peak-signal-to-noise ratio (PSNR) function:

$$PSNR=10\log_{10}\left[\frac{(peak-to-peak\ value\ of\ the\ original\ data)^2}{MSE}\right] \quad (1)$$

Where MSE, is the mean square error, defined as:

$$MSE=\frac{1}{NxM}\sum_{i=1}^{N}\sum_{j=1}^{M}\left[O(i,j)-R(i,j)\right]^2 \quad (2)$$

Where: $O\ (i,j)$ is the original frame, and $R\ (i,j)$ is the corresponding reconstructed frame of size NxM pixels, based on the obtained motion vectors. The error terms are not used in the frame reconstruction.

Critical PSNR

Let X(i,j,t) denote the reference frame and Yo(i,j,t) is the original current frame, both of them are corrupted by noise. It is well known that the covariance of the image is given by

$$E[X(i,j,t)X(i+u,j+v,t+dt)]=\sigma^2 R(|u|,|v|) \quad (3)$$

Where E[.] denotes the expectation, σ^2 is the variance of X(i,j,t), and (u,v) is the new location of the pixels at any time t+dt in both direction i and j respectively.

Let Yr(i,j,t) be the estimates of Yo(i,j,t) and let (du,dv) is the error of motion vector estimation (u,v) due to the noise. Thus, the mean square error between the original current frame and the reconstructed current frame is defined as

$$MSE=E\{[Yo(i,j,t)-Yr(i,j,t)]^2\} \quad (4)$$

But the original current frame Yo(i,j,t) and the reconstructed current frame Yr(i,j,t) can be written as

$$Yo(i,j,t)=X(i+u,j+v,t) \quad (5)$$

$$Yr(i,j,t)=X(i+u',j+v',t) \quad (6)$$

Let du and dv be the estimate error of u and v, respectively, due to the influence of noise then

$$du=u-u', \qquad dv=v-v' \quad (7)$$

Assuming du, dv to independent random variables thus, for small image variance relative to the noise variance equation (4) can be simplified using equation (3),(5),(6), and (7) to yield

$$MSE=2\sigma_n^2[1-E[R(|du|,|dv|)]] \quad (8)$$

Where σ_n^2 is the noise variance. For small du and dv equation (8) can be written as

$$MSE=2\sigma_n^2[1-R(E|du|,E|dv|)] \quad (9)$$

Equation (9) is useful for determined the critical PSNR, let

$$R(|p|,|q|)=\rho^{\sqrt{p^2+q^2}} \quad (10)$$

Thus, from equation (9), and (10) the mean square error between the original frame and the reconstructed frame is given by

$$MSE=2\sigma_n^2\left[1-\rho^{\sqrt{du^2+dv^2}}\right] \quad (11)$$

For good estimation typical value of ρ is around *0.95*. For example, assuming that the noise variance is *800*, ρ = 0.95, and the peak-to-peak value of the original frame is *255* (8-bit/pixel). Thus, if the estimation error is one block in one direction only, du = 1 and dv= 0, the PSNR is equal to *29.1*dB. If the error is one block in both direction, du = 1 and dv=1, the PSNR would be *27.6* dB. Thus, the PSNR is reduced by 1.5 dB due to the estimation error within one block. This error can be accepted so, the value of 1.5 dB with the subjective test is used as a guideline to determine the critical SNR. Three viewers judged the test picture with high-resolution monitors. Displaying the reconstructed and the original pictures conducted the evaluation.

3. Performance Evaluation

Three of famous video sequences have been used to evaluate the critical SNR. Each video sequence consists of different kinds of motions such as zooming, panning, fast object translation, and etc. They are summarized in table (1).

Sequences	Format	Kind of motion
Tennis	*352x240*	Zooming and panning
Carphone	*176x144*	Fast object translation
Foreman	*176x144*	Panning with object translation

Table 1. Different kinds of motion with the test video sequences

Motion estimation is performed on the luminance component of these video sequences with and without noise effect. Four motion estimation search algorithms (FS, TSS, NTSS, 4SS) were applied to eight frames of these video sequences. The motion vectors were found by using the mean absolute error (MAE) between blocks, and the PSNR were calculated in each case.

An additive Gaussian noise with different SNR, zero mean, and the variance is *800* was degraded the "Tennis" sequence. We have applied all motion estimation algorithms to this sequence. The results show that, the critical SNR for full search algorithm (FS) is *11*dB, three-step search algorithm (TSS) is *10* dB, the new three-step search algorithm (NTSS) is *10* dB, and four-step search algorithm (4SS) is *9* dB. This means that the best performance for the "Tennis" sequence were obtained using four-step search algorithm, and the same performance were obtained using three-step search algorithm and the new three-step search algorithm. The performance of full search algorithm is the worst performance in this case.

The results of applying the current motion estimation algorithms to the "Carphone" sequence corrupted by Gaussian noise, with zero mean and *800* variance, show that the critical SNR for all algorithms are the same and is equal to *25* dB. This result gives us the fact that, in case of fast object translation sequences the behavior of the PSNR for different value of SNR is the same and relatively high.

Using Gaussian noise with the same mean and variance to evaluate the critical SNR for the "Foreman" sequence. This process results in small critical SNR compared with the previous sequences. The results show that, the critical SNR for full search algorithm (FS) is *4*. dB, three-step search algorithm (TSS) is *4* dB, the new three-step search algorithm (NTSS) is *4.5* dB, and four-step search algorithm (4SS) is *4* dB.

The PSNR for each sequence vs. the frame number are shown in Figure 1 to Figure 3 respectively.

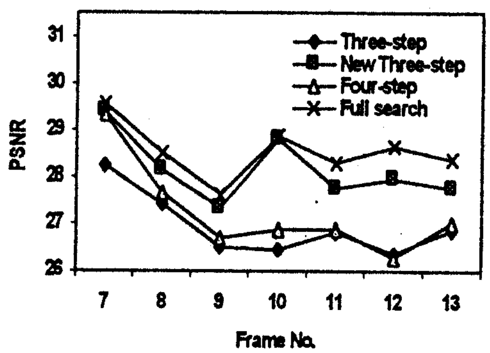


Fig.1 "Tennis" sequence: The PSNR performance for various search algorithms.

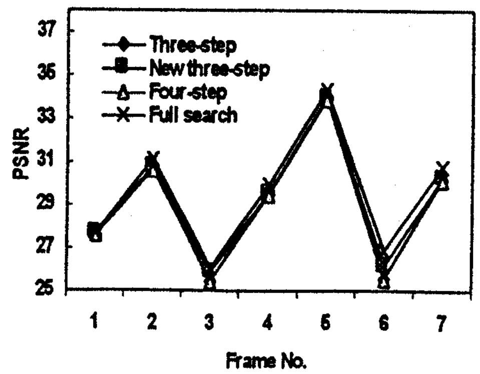


Fig. 2 "Carphone" sequence: The PSNR performance for various search algorithms.

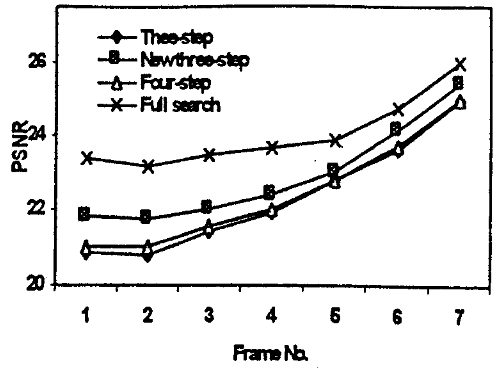
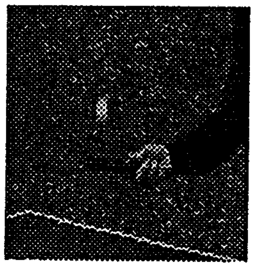
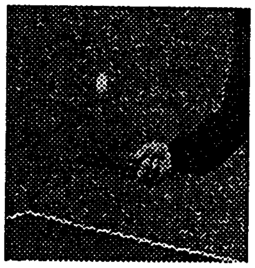


Fig. 3 "Foreman" sequence: The PSNR performance for various search algorithms.

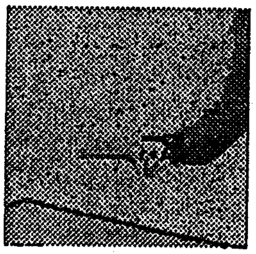
Figure 4 shows one of the estimated frame from the "Tennis" sequence with influence of zero mean Gaussian noise with SNR 10 dB, and noise variance 800.



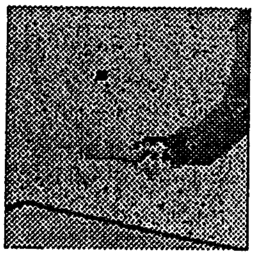
Original frame no.13,



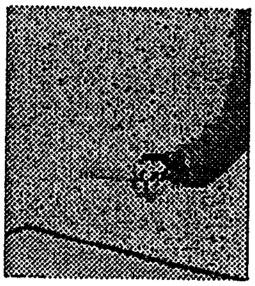
Original frame no.14,



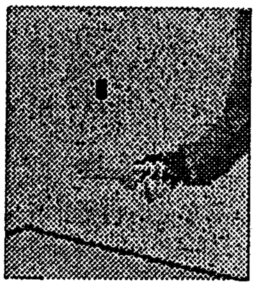
Reconstructed using TSS,



Reconstructed using TSS,



Reconstructed using 4SS,



Reconstructed using FS.

Fig. 4 "Tennis" reconstructed frame no.14 with influence of Gaussian noise(SNR = 10 dB, variance = 800)

Figure 5 to Figure 7 show the behavior of PSNR with respect to the SNR for two successive frames of the current sequences at noise variance 800.

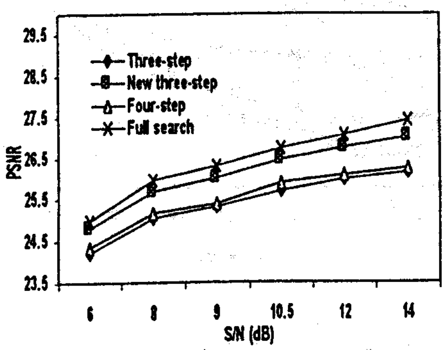


Fig. 5 "Tennis" frame 13 → 14: PSNR for various search algorithms.

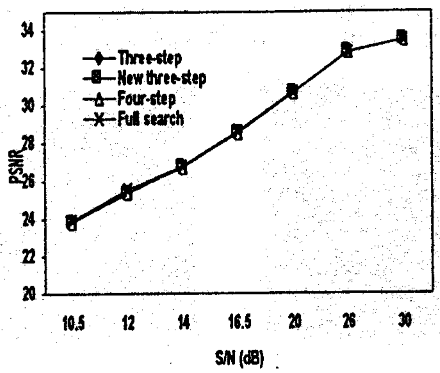


Fig. 6 "Carphone" frame 5 → 6: PSNR for various search algorithms.

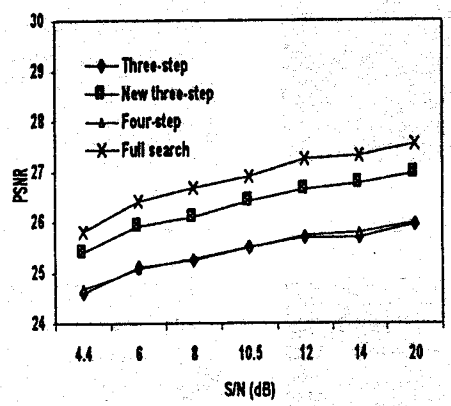


Fig. 7 "Foreman" frame 7 → 8: PSNR for various search algorithms.

4. Performance Evaluation with Variable Noise Variance

Performance evaluations of the current algorithms relative to the noise variance are obtained with 8x8-block size. The PSNR, with respect to the noise variance for the same frames in Figure 5 to Figure 7, are shown in Figure 8 to Figure 10.

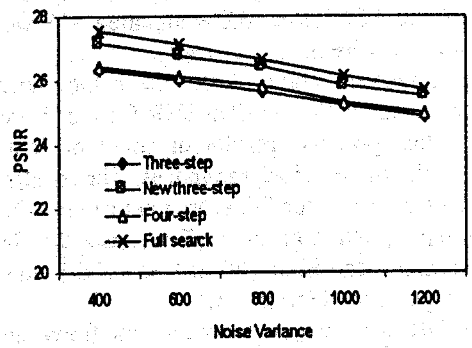


Fig. 8 "Tennis" frame 13 → 14: PSNR for various search algorithms (SNR = 10 dB).

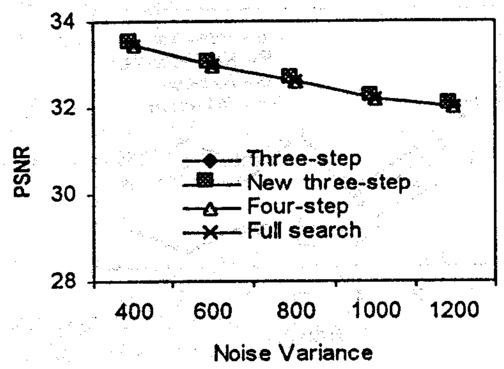


Fig. 9 "Carphone" frame 5 → 6: PSNR for various search algorithms (SNR = 25 dB).

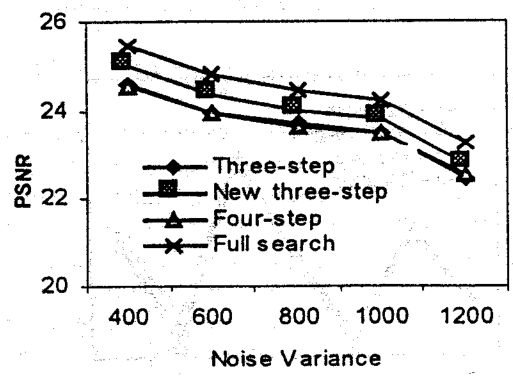


Fig. 10 "Foreman" frame 7 → 8: PSNR for various search algorithms (SNR = 5 dB).

5. Impact of block size

We applied all the search algorithms to the current sequences, for two different value of block size: 8 and 16. The MAE cost function is used for both algorithms. The obtained results of PSNR performance using full search and three-step search algorithms, without noise effect, for "Tennis" sequence, as an example, are shown in Figure 11. It can be concluded that, in all cases there is a very little change in the PSNR performance, when the block size was increased from 8 to 16.

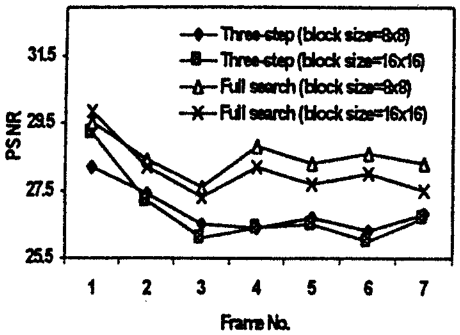


Fig. 11 "Tennis" sequences: PSNR performance of full search algorithm, and three-step search algorithm for two different block size 8x8 and block size 16x16.

Figure 12 show the behavior of the PSNR with respect to the SNR for frame 13→14 of "Tennis" sequence using block size 16x16 at noise variance 800. From this Figure it is clear that, The critical SNR value is approximately equal to the same value for block size 8x8.

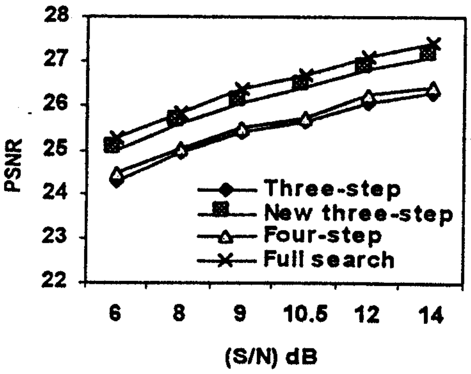


Fig. 12 "Tennis" frame 13→14:PSNR for various algorithms With block size 16x16, and noise variance 800.

6. Conclusion

In this paper, we have presented a performance evaluation of three fast motion estimation algorithms (TSS, NTSS, 4SS) and the full search algorithm (FS) with influence of zero-mean Gaussian noise. The results of applying these algorithms to a video sequence that contain zooming and panning are presented which show 9-11 dB SNR are the minimum limit for good estimation. For video sequences of panning and fast object translation, small value of SNR is accepted and the same value for all algorithms (from 4-5 dB). It is obvious that, for video sequences of fast object translation the critical SNR (*25* dB) is relatively higher than the other video sequences. Moreover, it is observed that the worst performance is often in case of full search algorithm. The probability of error is increased when the number of checkpoint within the search window is increased. PSNR decreased slowly and linearly as the noise variance increased at the value of critical SNR. As we noted from equation (11) for small error in motion estimation, the PSNR is linearly changed with noise variance at constant ρ. Also, it is obvious that, with all the search algorithms used in this paper, it is better to select smaller block size for the current video sequences which will reduce the complexity of the search algorithm and produce the same performance with present of zero mean Gaussian noise.

References

[1] F. Giorda and A. Racciu, "Bandwidth reduction of video signals via shift vector transmission," IEEE Trans. Comm., vol. COM-23, pp.1002-1004, Sept. 1975.

[2] A. N. Netravali and J. D. Robbins," Motion-compensated television coding Part I," Bell Syst. Tech. J., vol.58, pp. 631-669, Mar. 1979.

[3] T. Kago, K. Iinuma, A. Hirano, Y. Iijima, and T. Ishiguro, " Motion compensated interframe coding for video conferencing," in Proc. Nat: Telecommun. Conf., New Orleans, LA, pp. G5.3.1-5.3.5., Nov.29-Dec.3, 1981.

[4] Renxiang Li, Bing Zeng, and Ming L.Liou, " A new three-step search algorithm for block motion estimation," IEEE Trans. Circuits and systems for video Technology, vol.4, No.4, pp. 438-442, August 1994.

[5] Lai-Man Po and Wing-Chung Ma," A novel four-step search algorithm for fast block motion estimation," IEEE Trans. Circuits and systems for video technology, vol.6, No.3, pp. 313-317, Jun. 1996.

[6] M.J. Chen, L.G. Chen, and T.D. Chiueh," One dimensional full-search motion estimation algorithm for video coding," IEEE Trans. Circuits and systems for video Technology, vol.4, pp. 504-509, Oct.1994.

[7] M. Ghanbari, " the cross-search algorithm for motion estimation," IEEE Trans. Comm., vol.38, No.7, pp. 950-953, July 1990.

[8] B. Liu, and A. Zaccarin," New fast algorithms for the estimation of block motion vectors," IEEE Trans. Circuits & systems, vol.3, No.3, pp.148-157, April 1993.

[9] K. Xie, L.V. Eycken, and A. Oostelinck, " A new block-based motion estimation algorithm," Signal Processing: Image Communication, vol. 4, pp. 507-517, May 1992.

[10] L.W. Lee, J.F. Wang, J.Y. Lee, and J. D. Shie, " Dynamic search-window adjustment and interlaced search for block-matching algorithm," IEEE Trans. Circuits and systems for video Tech., vol.3, pp. 85-87, Feb. 1993.

[11] J. R. Jain, and A. K. Jain," Displacement measurement and its application in interframe image coding," IEEE Trans. Commun., vol. Com-29, pp. 1799-1808, Dec. 1981.

About authors...

Sayed USAMA was born in BENI-SOUEF (EGYPT) in 1964. He received the B.Sc, and M.SC. degree at Electrical Engineering Department in ASSIUT University, EGYPT. Currently he is a postgraduate student in Telecommunication Department at CTU in PRAGUE, Faculty of Electrical Engineering. His field of study is in the Motion Estimation of Video Sequences.

Boris ŠIMÁK was born in 1956 in Hradec Králové, Czech Republic. He received the Ing. (M.Sc.) and C.Sc (Ph.D.) degree in field of Telecommunication Engineering from the CTU in Prague, FEE in 1981 and 1987 respectively. From 1996 he worked as a Doc. (Associate Professor) at FEE, CTU in Prague. At present time he is a head of the Department of Telecommunication Engineering on CTU in Prague, FEE. His research interest is in Digital Signal Processing.

Miloš Klíma graduated at the CTU in 1974 (1974 - MSEE, 1978 - PhD, 1991 - Assoc.Prof.). He is employed by the CTU from 1974 in the positions as a Researcher, Assistant Professor and Associate Professor. His field of interest is Television Technology, Image Processing and Photonics. His research activity is focused on optical information processing, image processing and image sensing.