

Design & Implementation of Motion Detection & Reconstruction of Image Using Modified Pixel Motion Approach

Navneet Kaur¹ Amit Doegar²
^{1,2}Student

^{1,2}Department of Computer Science & Engineering

^{1,2}National Institute of Technical Teachers Training and Research, Sector 26, Chandigarh, India

Abstract— The human perception has the capability to acquire, integrate, and interpret all this abundant visual information around us. It is challenging to impart such capabilities to a machine in order to interpret the visual information embedded in still images, graphics, and video or moving images in our sensory world. Disparate as they may seem, these many applications share one common thread in all such applications, the demand is high for accurate estimates of motion requiring minimal computational cost. In this, a new technique called estimating the difference between two frames of image using Pixel motion estimation technique is proposed. The proposed technique attempts to decrease the computation time between the current frame and the next frame. The displacement of each picture element in each frame forms the displacement vector field (DVF) and its estimation can be done using at least two successive frames. The pixel based approaches depend upon intensity of image and its performance is affected by presence of noise. In this, it also proposes real time motion estimation and enhancement based difference of frames. The PSNR and MSE are also calculated. All simulations are done in MATLAB 8.1.

Key words: motion detection, motion estimation, Pixel Motion Detection, Enhancement based detection

I. INTRODUCTION

The field of digital image processing refers to processing digital images by means of a digital computer. Note that a digital image is composed of a finite number of elements, each of which has a specific location and value. These elements are denoted to as picture elements, image elements, pels, and pixels. Pixel is the term widely used to signify the elements of a digital image. Vision is the most advanced of our senses, so it is not surprising that images play the single most important role in human perception. However, unlike humans, who are limited to the visual band of the electromagnetic spectrum, imaging machines cover almost the entire EM spectrum, ranging from gamma to radio waves. They can operate on images generated by sources that humans are not accustomed to associating with images [1].

One of the first applications of digital images was in the newspaper industry, when pictures were first sent by submarine cable between London and New York. Introduction of Bartlane cable picture transmission system in early 1920s reduced the time required to transport a picture across the Atlantic from more than a week to less than three hours. Specialized printing apparatus coded pictures for cable transmission and then reconstructed them at the receiving end. The early Bartlane systems were capable of coding images in five distinct levels of gray. This capability was increased to 15 levels in 1929 [2].

In parallel with space applications, digital image processing techniques began in late 1960s and early 1970s

to be used in medical imaging, remote Earth resources observations, and astronomy. The invention in early 1970s of computerized axial tomography (CAT), also called computerized tomography (CT) for short, is one of the most important events in the application of image processing in medical analysis. Computerized axial tomography is a process in which a ring of detectors encircles an object (or patient) and an X-ray source, concentric with detector ring, rotates about the object. The X-rays pass through the object and are collected at the opposite end by the corresponding detectors in the ring [2].

In the design and analysis of image processing systems, it is suitable and often necessary mathematically to characterize the image to be processed. There are two basic mathematical characterizations of interest: deterministic and statistical. Indeterministic image representation, a mathematical image function is defined and point properties of the image are considered. The first step towards designing an image analysis system is digital image acquisition using sensors in optical or thermal wavelengths. A two-dimensional image that is recorded by these sensors is drawing of the three-dimensional visual world. The captured two dimensional signals are sampled and quantized to yield digital images [3].

Another aspect of image processing involves compression and coding of the visual information. With growing demand of various imaging applications, storage requirements of digital imagery are growing explosively. Digital video is now an integral part of many aspects of business, education and entertainment, from digital TV to web-based video news. Digital video is visual information represented in a discrete form, suitable for digital electronic storage and/or transmission. The goal of a video coding system is to support video communications with an 'acceptable' visual quality: this depends on the viewer's perception of visual information, which in turn is governed by the behavior of the human visual system. Measuring and quantifying visual quality is a difficult problem. Compact representation of image data and their storage and transmission through communication bandwidth is a crucial and active area of development today [4].

Today, there is almost no area of technical endeavor that is not impacted in some way by digital image processing. The areas of application of digital image processing are so varied that some form of organization is desirable in attempting to capture the breadth of this field. One of the simplest ways to develop a basic understanding of the extent of image processing applications is to categorize images according to their source [5].

The paper is ordered as follows. In section II, we discuss about video processing & its system. In Section III, It defines basic motion estimation & compensation system and its technique. In Section IV, it describes proposed technique of motion estimation. After this, it designates the

main results of this system. Finally, conclusion is explained in Section VI.

II. VIDEO PROCESSING

A video image is a projection of a 3-D scene onto a 2-D plane (Figure 1). A 3-D scene consisting of a number of objects each with depth, texture and illumination is projected onto a plane to form a 2-D representation of the scene. The 2-D representation contains varying texture and illumination but no depth information. A still image is a 'snapshot' of the 2-D representation at a particular instant in time whereas a video sequence represents the scene over a period of time. A 'real' visual scene is continuous both spatially and temporally. In order to represent and process a visual scene digitally it is necessary to sample the real scene spatially (typically on a rectangular grid in the video image plane) and temporally (typically as a series of 'still' images or frames sampled at regular intervals in time) as shown in Figure 1. Digital video is the representation of a spatio-temporally sampled video scene in digital form. Each spatio-temporal sample is represented digitally as one or more numbers that describe the brightness (luminance) and colour of the sample [6].

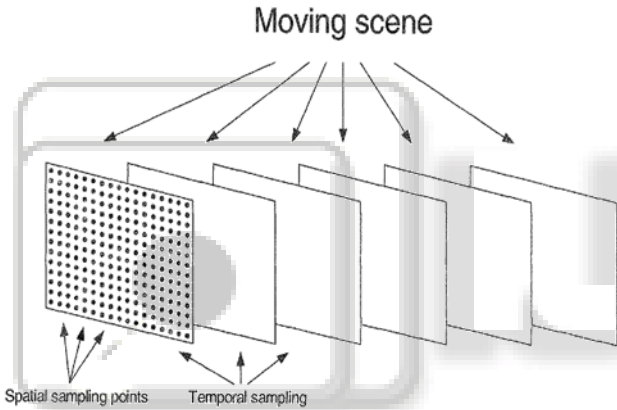


Fig. 1: Spatial and Temporal Sampling [2]

A digital video system is shown in Figure 2. At the input to the system, a 'red' visual scene is captured, typically with a camera and converted to a sampled digital representation. This signal may be handled in number of ways include processing, storage and transmission. At output of system, video is displayed by reproducing 2-D video image on 2-D display [7].

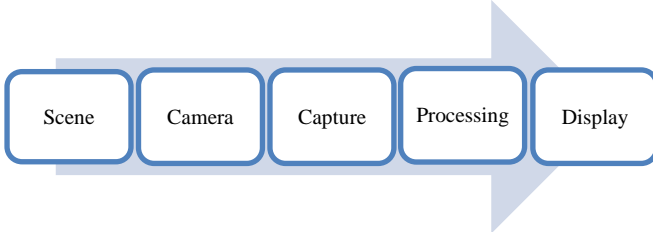


Fig. 2: Digital Video System [2]

Video is captured using a camera or a system of cameras. Most digital video systems use 2-D video captured with a single camera. The camera focuses a 2-D projection of video scene on to a sensor. In case of color image capture, each color component is filtered and projected. Generating a digital representation of a video scene can be considered in two stages: acquisition (converting a projection of the scene into an electrical signal) and

digitization (sampling the projection spatially and temporally) [8].

A digital image may be generated by sampling an analog video signal (i.e. varying electrical signal that presents a video image) at regular intervals. The result is a sampled version of the image: the sampled image is only defined at a series of regularly spaced sampling points. The most common format for a sampled image is a rectangle. A moving video image is formed by sampling the video signal temporally, taking a rectangular 'snapshot' of the signal at periodic time intervals. Playing back the series of frames produces the illusion of motion. A higher temporal sampling rate called frame rate gives a 'smoother' appearance to motion in the video scene but requires more samples to be captured and stored [9].

| Video Frame Rate | Appearance |
|------------------|--------------------------|
| <10 fps | Unnatural Appearance |
| 10-20 fps | Slow Movement appears Ok |
| 20-30 fps | Movement is smooth |

Table 1: Video Frame Rates

Frame rates below 10 frames per second are sometimes used for very low bit-rate video communications (because the amount of data is relatively small): however, motion is clearly jerky and unnatural at this rate. Between 10 to 20 frames per sec is more typical for low bit communications. 30 fps is standard for television pictures (Refer Table 1). The visual appearance of a temporally sampled video sequence can be improved by using interlaced video used for broadcast quality television signals. Displaying a 2-D video signal involves recreating each frame of video on a 2-D display device. The most common type of display is the cathode ray tube (CRT) in which the image is formed by scanning a modulated beam of electrons across a screen [10].

A monochrome video image may be represented using one number per sample. This number indicates brightness of each sample and a large number indicates a brighter sample. If a sample is represented using 8 bits then number 0 represents black color while a value 255 represents white color. Luminance is commonly represented with 8 bits per sample. Higher luminance is used for specialist applications [11].

III. MOTION ESTIMATION & COMPENSATION

Motion estimation creates a model of the current frame based on available data in one or more previously encoded frames ('reference frames'). These reference frames may be 'past' frames (i.e. earlier than the current frame in temporal order) or 'future' frames (i.e. later in temporal order). The design goals for a motion estimation algorithm are to model the current frame as accurately as possible (since this gives better compression performance) while maintaining acceptable complexity. Motion estimation (ME) techniques have been successfully useful in motion compensated predictive coding for falling temporal redundancies. An effective representation of motion is serious in order to reach high performance in video coding. Estimation techniques should on one hand provide good prediction [12].

A video sequence can be considered to be a discretized three-dimensional prediction of real four-dimensional continuous space-time. The objects in real world may move, rotate, or deform. The movements cannot be detected directly, but instead the light reflected from the object surfaces and estimated onto an image. The light source can be moving, and the reflected light varies depending on the angle between a surface and a light source. A perfect motion model would take all the factors into account and find motion that has the maximum likelihood from observed video sequence [10].

In figure 3, the motion estimation module creates a model by modifying one or more reference frames to match the current frame as closely as possible (according to a matching criterion). The current frame is motion compensated by subtracting the model from the frame to produce a motion-compensated residual frame. This is coded and transmitted, along with the information required for the decoder to recreate the model (typically a set of motion vectors). At the same time, the encoded residual is decoded and added to the model to reconstruct a decoded copy of the current Frame (which may not be identical to the original frame because of coding losses). This reconstructed frame is stored to be used as a reference frame for further predictions [13].

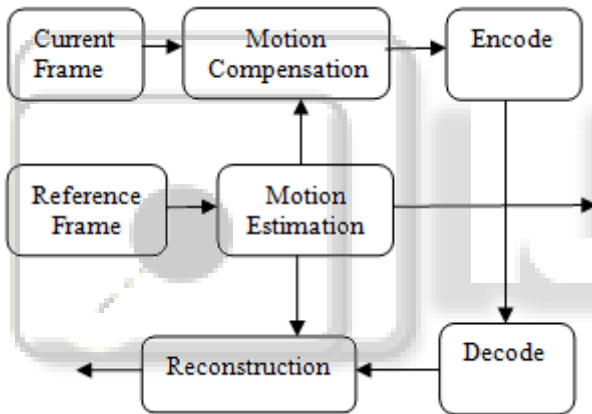


Fig. 3: Motion Estimation & Compensation

Changes between frames are mainly due to movement of objects. Using a model of motion of objects among frames, the encoder approximates motion that happened between reference frame and current frame. This process is called motion estimation (ME) [4]. The encoder then uses this motion and information to move contents of reference frame to deliver a better prediction of current frame. This process is known as motion compensation (MC), and prediction so produced is called motion-compensated prediction (MCP) or the displaced-frame (DF). In this case, coded prediction error signal is called displaced-frame difference (DFD) [14].

Block-matching motion estimation is extensively used motion estimation method for video coding. The process then undertakes that all pels within the block undergo same translational movement. This motion vector is projected by searching for best match block in a larger search window pels centred at same location in a reference frame [15]. This procedure is based on a translational model of motion of objects between frames. Inter frame predictive coding is used to remove the huge expanse of temporal and spatial redundancy that happens in video sequences and

helps in compressing them. The better the prediction, smaller the error and hence the transmission bit rate when there is motion in a sequence, then a pel on same part of the moving object is a better prediction for the current pel. The another important parameter is block size. If the block size is smaller, it attains better prediction quality. This is due to a numeral of reasons. A smaller block size decreases effect of accuracy problem. In addition, a smaller block size delivers a better piecewise translational approximation to non-translational motion.

IV. PIXEL MOTION ESTIMATION AND RECONSTRUCTION

Vision is the most advanced of our senses, so it is not surprising that images play the single most important role in human perception. Motion estimation (ME) methods have been successfully useful in motion compensated analytical coding for reducing temporal redundancies. They belong to the class of nonlinear predictive coding systems. An efficient representation of motion is critical in order to reach high performance in video coding.

A. Real Time Pixel Motion Estimation On Videos:

For this, it requires the interfacing of webcam with computer. Making the interface between computer and the camera can be done using lower level language like C or C++ will give lots of elasticity, but it will also appeal lots of work and background knowledge. We use MATLAB for this purpose. In this, it depends upon forward prediction coding. In this, we use the frame rate of 30 fps with RGB image. Initially set the large buffer size for frames storage and also set the triggering manually.

Forward prediction involves using an 'older' encoded frame (i.e. a preceding frame in temporal order) as prediction reference for the current frame. Forward prediction performs poorly in certain cases, for example: when there is a significant time difference between the reference frame and the current frame (which may mean that the image has changed significantly); or when a scene change or 'cut' occurs.

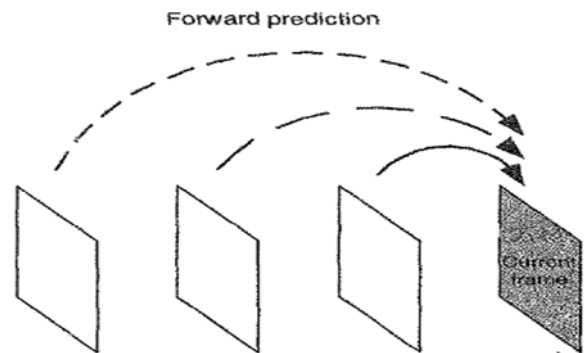


Fig. 4: Forward Predictions [2]

1) Algorithm:

- (1) Set the camera settings manually.
- (2) Start the camera using MATLAB tool.
- (3) Set the triggering so that it takes frames continuously.
- (4) Apply required operations on these frames.
- (5) Extract the main features of these frames.
- (6) Detect the final motion and also locate the position of motion estimation.

- (7) Calculate performance parameters like MSE, PSNR and computation time.
- (8) Reconstruct the frame by using block based concept.

B. Motion Estimation & Reconstruction Of Images:

In this case, a primary image will be taken from webcam and converted to black and white image. The image will then be accessed and using different edge detection techniques, the required area will be segmented out. A filter is used to remove noise. Such noise reduction is a typical pre-processing step to improve the results of later processing. The pixel subtraction operator takes two images as input and produces as output a third image whose pixel values are simply those of the first image minus the corresponding pixel values from the second image. A common use is to subtract background variations in illumination from a scene so that the foreground objects in it may be more easily analyzed. At last, reconstruction process is applied on detected output so that original frame is produced.

C. Motion Estimation With Enhancement:

Enhancement refers to accentuation or sharpening of image features, such as contrast, boundaries, edges, etc. The process of image enhancement, however, in no way increases the information content of the image data. It increases the dynamic range of the chosen features with the final aim of improving the image quality. An outdoor scene captured on a foggy winter morning could result in a blurred image. In this case the degradation is due to the fog and mist in the atmosphere, and this type of degradation is known as atmospheric degradation. In some other cases there may be a relative motion between the object and the camera. Thus if the camera is given an impulsive displacement during the image capturing interval while the object is static, the resulting image will invariably be blurred and noisy. Thus it may be observed that in all the above cases the resultant image is a degraded version of the original object. For this, enhancement is required. So, in proposed method, the input frames of images are enhanced initially and then different operations are applied on these frames so that necessary motion estimation and reconstruction is carried out.

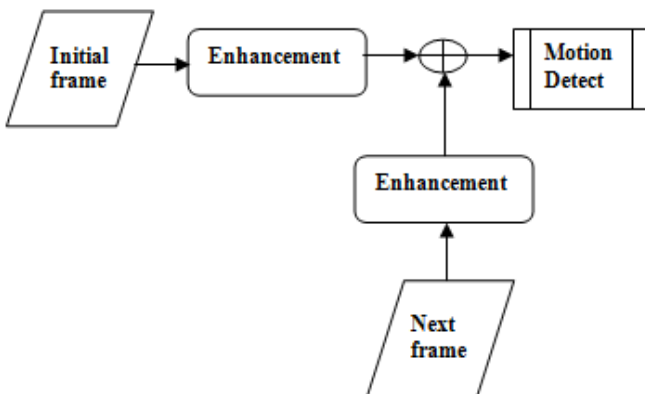


Fig. 5: Designing Steps of Motion Estimation with Enhancement

V. PERFORMANCE RESULTS & ANALYSIS

In this, it provides the main results of proposed techniques used for motion estimation & reconstruction. The output images contains real time videos output (shown in fig 6), real time still image motion estimation output (figure 7 & 8) and enhancement based motion output (shown in fig 9 & 10). All results are implemented in MATLAB 8.1. The main objective of this work is to reduce the computation time to detect motion between consecutive frames. So, in this, it works on real time images taken by webcam. It provides compressed data. Then various operations are applied on these frames to get final motion and then reconstruction process is applied to get original image output.

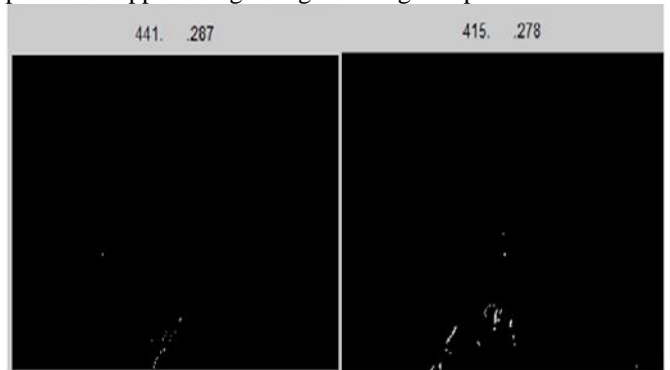


Fig. 6: Real Time Compressed Video Motion Output

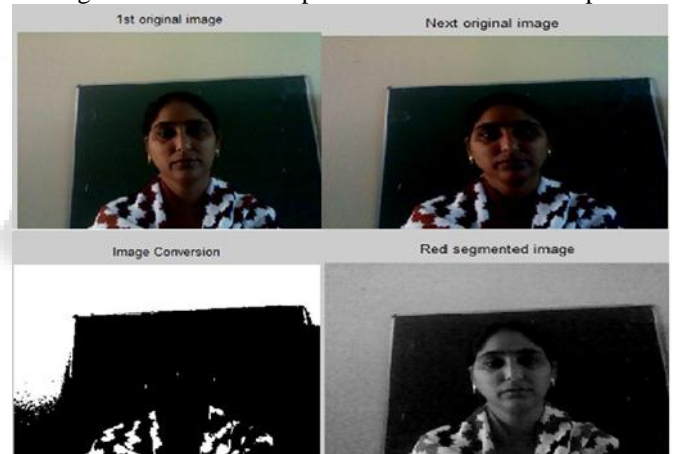


Fig. 7: Real Time Compressed Image Operations Output



Fig. 8: Reconstructed Compressed Motion Image Output



Fig. 9: Enhancement Based Motion Image Output



Fig. 10: Enhancement Based Reconstructed Motion Image Output

The performance of proposed approach is shown in Table 2. This table compares this by performance parameters PSNR and computation time.

| Parameters | Proposed Approach |
|-------------------------------------|-------------------|
| PSNR | 21 |
| RMSE | 508 |
| Computation Time | 5 sec |
| Performance at Delay=0.1 sec | |
| PSNR | 27 |
| RMSE | 126 |
| Computation Time | 4 sec |
| Performance at Delay=1 sec | |
| PSNR | 21 |
| RMSE | 508 |
| Computation Time | 5 sec |

Table 2: Parameters Analysis

VI. CONCLUSION

In this work, it proposed motion estimation using modified pixel based approach while reconstruction is using block based approach so that our computation time gets decreased. It also proposes motion estimation with enhancement so that it adjusts contrast of input frames. It also proposes real time compressed video motion estimation. Motion Estimation forms a major computation bottleneck in video processing applications such as the detection of noise in image sequences, interpolation/ prediction of missing data in image sequences and de-interlacing of image sequences. The proposed algorithm can reduce the computational time as compared to block based technique. The pixel based approaches depend upon intensity of image and its performance is affected by presence of noise. While block based techniques depends upon motion vectors and it has high computation time as compared to pixel approaches parameters. The MSE and PSNR value of proposed method is also calculated.

REFERENCES

- [1] R. Gonzalez, R. Woods and S. Eddins "Digital Image Processing Using Matlab", 2004, Prentice Hall.
- [2] Gonzalez, R., Woods, R., "Digital Image Processing", 2nd Edition, Prentice-Hall, (2002).
- [3] Acharya, T. and Ray, A.K, Image processing: Principles and Application, Wiley, and Hoboken, NJ, 2005.
- [4] E. G. Richardson. H.264 and MPEG-4 video compression. Wiley, Chichester, England, 2003.
- [5] C. Zhu, X. Lin, L. Chau, "Hexagon-Based Search Pattern For Fast Block Motion Estimation", IEEE Transactions On Circuits And Systems For Video Technology, Vol. 12, No. 5, May 2002
- [6] C.H. Cheung, L.M. Po, "Novel Cross-Diamond-Hexagonal Search Algorithms for Fast Block Motion Estimation", IEEE Transactions On Multimedia, Vol. 7, No. 1, February 2005.
- [7] M. Gallant, F. Kossentini, "An Efficient Computation-Constrained Block-Based Motion Estimation Algorithm for Low Bit Rate Video Coding".
- [8] Ahmad, W. Zheng, J. Luo, M. Liou, "Fast Adaptive Motion Estimation Algorithm", IEEE Transactions On Circuits And Systems For Video Technology, Vol. 16, No. 3, March 2006.
- [9] K. Ng, L. Po, K.M. Wong, "Search Patterns Switching For Motion Estimation Using Rate of Error Descent", ICME 2007.
- [10] S. Goel, M. A. Bayoumi, "Multi-Path Search Algorithm for Block-Based Motion Estimation", ICIP 2006.
- [11] M. Tourapis, "Enhanced Predictive Zonal Search For Single And Multiple Frame Motion Estimation".
- [12] D. Kesrarat, V. Patanavijit, "Empirical Study on Performance Comparisons of Block-Based Motion Estimation on Multi Sub-Pixel Displacement with Multiples Block Size", 26th International

- Conference on Advanced Information Networking and Applications Workshops, 2012.
- [13] S. Klomp, M. Munderloh, Y. Vatis, J. Ostermann, "Decoder-Side Block Motion Estimation For H.264 / Mpeg-4 Avc Based Video Coding" IEEE Institute For Informations, 30167 Hannover, Germany 2009.
- [14] L.M. Po, C. W. Ting, K.M. Wong, K.H. Ng, "Novel Point-Oriented Inner Searches For Fast Block Motion Estimation", IEEE Transactions On Multimedia, Vol. 9, No. 1, January 2007.
- [15] L.M. Po, K.H. Ng, K.M. Wong, K.W. Cheung, "Multi-Direction Search Algorithm for Block-Based Motion Estimation" Department Of Electronic Engineering, City University Of Hong Kong 2008.

