CHAPTER 2

LITERATURE SURVEY

2.1 INTRODUCTION

The literature review undertaken as a part of the thesis is focused on understanding the basics of video compression, the importance of shot detection and key frame extraction in video analysis, different intraframe compression techniques and the methods available for reducing the temporal correlation between the adjacent frames in a video and application of these techniques to compress patient video signal for telemetry applications.

The first section focuses on the literature survey related to the detections of shot in a video sequence. Following this, different techniques for extracting key frames from a video sequence are discussed.

2.2 VIDEO SHOT DETECTION

The advances in digital video technology and the ever increasing availability of computing resources have resulted in the last few years in an explosion of digital video data, especially on the Internet. In any digital video processing operations like indexing, retrieval and compression, the preliminary step is the identification of shot boundaries in a video. Several reviews on shot boundary detection have been published in the last decade and a few of the literatures relating to this dissertation are listed below.
Zhang et al (1993) have proposed a pixel based difference method for shot boundary detection. Although slow, it produced good results once the threshold was manually tailored to the video sequence. The change between the two frames can be detected by comparing the differences in intensity values of corresponding pixels in two frames. Feature frame matrix for each frame was created and the distance between successive frames was calculated. When this distance exceeded a threshold, shot boundary was detected.

Gong and Liu (2001) have proposed a video summarization method that minimizes the visual content redundancy from the input data. A visual content redundancy metric was derived and it indicates how much the video can be curtailed without losing too much visual content. For shot boundary detection, singular value decomposition method has been used.

Ullas Gargi et al (2000) have made a performance evaluation and characterization of a number of shot-change detection methods that use color histograms, pixel difference, and block motion matching on video data. Threshold selection is important in video boundary detection and local window averaging works well.

Ferman et al (2002) have presented various types of histogram based description for shot detection. The difference between the histogram of successive frames was computed and large frame differences were marked as possible shot changes. Comparison of histograms was done based on Chi-square comparison, absolute difference between corresponding bins and histogram intersection.

Wei Jyh Heng and Ngan (2002) have explained a novel shot boundary refinement algorithm which works independently of the contents and transition length. The algorithm is accurate under most type of transitions
and it detects shot boundary based on the similarity of a reference with frame multiple frames.

Seong-whan Lee (2000) has proposed a fast scene change detection algorithm using direct feature extraction from MPEG compressed videos. Binary edge maps are derived from the ac coefficients in blocks which were discrete cosine transformed. Then edge orientation, strength and offset are measured using correlation between the ac coefficients in the derived binary edge maps.

Wang et al (2004) have proposed an alternative complementary framework for quality assessment based on the degradation of structural information. They developed a measure of Structural SIMilarity (SSIM) that compares local patterns of pixel intensities that have been normalized for luminance and contrast.

Huang and Liau (2001) have proposed a robust technique based on pixel and histogram differences for video segmentation. The combination of intensity and motion information is used to detect the abrupt and gradual scene changes. They have also developed a novel intensity statistics for detecting gradual scene changes.

2.3 KEY FRAME EXTRACTION

The efficiency of video compression process lies on the identification of key frames followed by intra and interframe prediction coding. This section briefly describes the different key frame algorithms reported in the literature.
Tonomura et al (1993) have described a video abstraction where the first frame in a shot is treated as key frame. Simple technique but the first frame may not be a good abstraction of the entire shot.

In Rui et al (1999), video is time sampled regardless of shot boundaries, first and last frame are treated as key frames.

Hanjalic et al (1997) have proposed a histogram based key frame extraction technique. Difference between consecutive frames in terms of color histogram for visual features is compared with certain threshold value. And the key is selected, if the value is greater than the threshold.

Xu-Dong Zhang et al (2003) have proposed a key frame extraction technique where the frames are grouped in clusters and the key frames are selected from the largest clusters.

In Gong et al (2000), the video frames are time sampled and visual features computed from them. The refined feature space obtained by the SVD is clustered, and a key frame is extracted from each cluster I-frame from the video and these I-frames are considered as key frames.

Video summarization, aimed at reducing the amount of data that must be examined in order to retrieve the information desired from information in a video, is an essential task in video analysis and indexing applications. Ciocca Gianluigi and Schettini Raimondo (2006) have proposed an innovative approach for the selection of representative (key) frames of a video sequence for video summarization. By analyzing the differences between two consecutive frames of a video sequence, the algorithm determines the complexity of the sequence in terms of changes in the visual content expressed by different frame descriptors. The algorithm, which escapes the complexity of existing methods based, for example, on clustering...
or optimization strategies, dynamically and rapidly selects a variable number of key frames within each sequence. The key frames are extracted by detecting curvature points within the curve of the cumulative frame differences.

### 2.4 INTRAFRAME COMPRESSION

Intraframe compression is similar to compression of still images and many algorithms have been developed for the compression of still images. JPEG (Joint Photographic Expert Group) was established as the first international standard for 2D image compression (Bhaskaran and Konstantinides 1996). It has four modes of operation namely, the sequential DCT based mode, the progressive DCT- based mode, the sequential lossless mode, and the hierarchical mode. The first two DCT based modes provide lossy compression whereas the last two are meant for lossless compression (Pennbaker and Mitchell 1993). In lossy JPEG compression, the image is split up into blocks, before processing using 2D DCT. As a result, the correlation across blocks is eliminated and blocking artifacts are introduced. They are undesirable and noticeable, especially at lower bit rates (Saha 2000). Other popular image compression techniques are the wavelet based Embedded Zero Trees (EZW) (Shapiro 1993) and Set Partitioning In Hierarchical Trees (SPIHT) coding (Said and Pearlman 1996). The wavelet transform uses sub-band coding where the image is passed through a low pass and high pass filter and then down sampled by a factor of two. If required, further level decompositions are done on the low pass filtered components.

Morris and Britch (2000) have described a baseline wavelet based video codec. It is designed to take a colour video sequence as input, and compress it in the manner adopted by MJPEG. Test sequences were compressed with the codec, and the results obtained indicate that the codec
outperforms MJPEG in terms of compression ratio, at the same level of quality.

A new coding technique called embedded block coding with optimized truncation (EBCOT) was introduced as an improvement to the wavelet coders (Lawson and Zhu 2002).

Secker and Taubman (2004) have proposed a highly scalable video compression scheme based on wavelet transforms. They have proposed two methods to determine the optimal trade-off between the motion and sub-band bit rates. They have explained the application of wavelet transforms to the motion vector fields and also to the embedded block coding.

Lei Zhanga et al (2005) have presented a simple and a fast coding technique for lossless compression of mosaic video data. The design of a video codec needs to strike a balance between the compression performance and the codec throughput. Aiming to make the encoding throughput high enough for real-time lossless video compression, the authors have proposed a hybrid scheme of inter and intraframe coding. Interframe predictive coding is invoked only when the motion between adjacent frames is modest and a simple motion compensation operation can significantly improve the compression performance. Otherwise, still frame compression is performed to keep the complexity low.

Meiqing Wang and Choi-Hong Lai (2005) have proposed a hybrid algorithm highlighting the advantages of the cube-based and the frame-based fractal compression methods. Experimental results show that the hybrid algorithm improves the compression ratio and the quality of decompressed images.
2.5 TEMPORAL REDUNDANCY

Ding Gui-guang and Guo Bao-long (2004) have proposed an algorithm to reduce the computational complexity using the line search strategy and a parallel search pattern. Only 9 search points, which is 4 search points less than diamond search algorithm are used in this algorithm. So the computational complexity is reduced.

Xuan Jing et al (2004) have proposed an efficient three-step search algorithm which employs a small diamond pattern in the first step, and an unrestricted search step is used to search the center area. Experimental results show that the new efficient three-step search performs better than three-step search in terms of mean square error and it requires less computation.

Zhi-Yi Mai et al (2006) have proposed an improved Motion Estimation Method Based on SSIM (MEBSS). Experimental results show that the MEBSS can reduce average bit rate and encoding time while maintaining the same perceptual video quality.

Cheung and Po (2002) have proposed a new algorithm using cross search pattern as the initial step and large/small diamond search patterns as the subsequent steps for fast block motion estimation. The algorithm uses the half way stop technique and finds small motion vectors with fewer search points than the diamond search algorithm.

Ying Li and Sayood (2007) have proposed an adaptive lossless video compression algorithm based on predictive coding. The algorithm exploits temporal and spatial redundancy in a backward adaptive manner with extremely low side-information.
Jang – Jer Tsai and Hsueh – Ming Hang (2009) have described a Pattern Based Block Motion Estimation (PBME) algorithm. A statistical PBME model which consists of two components has been described. Based on the model, a novel genetic rhombus pattern search algorithm has been proposed.

Hezerul et al (2004) have discussed the challenges of low rate video frame interpolation. Most existing algorithms interpolate at high frame rate. A multi resolution motion estimation algorithm has been explained. It estimates the movement first at lower resolution and then successively increases the resolution. It reduces the artifacts and long computation time.

Amit and Bawane (2009) have presented a low complexity novel cross diamond hexagonal search algorithm which uses two hexagonal search pattern in conjunction with diamond search pattern and it performs more efficiently than diamond search and hexagonal search algorithm with lower computational complexity and with the same quality as that of diamond search algorithm.

Avishek et al (2008) have described a boundary based approach towards pixel decimation with applications in Block Matching Algorithms (BMAs). The macroblocks are selected based on the boundary region matching only. The boundary based patterns are used to speed up motion estimation with marginal loss in image quality. New pixel decimation pattern has been described by combining the boundary based pattern and the genetic algorithm based pattern.

Anastasios Hamosfakidis and Yakup Paker (2002) have described a Hexagonal Search (HS) algorithm with center-biased checking point pattern for fast block motion estimation. The HS is compared with full search (FS), Four-Step Search (4SS), New Three-Step Search (NTSS), and Diamond
Search (DS) methods and the results show that the hexagonal search algorithm provides competitive performance with reduced computational complexity compared with other fast BMAS and FS algorithms and also provides similar performance to FS in terms of PSNR.

Memon and Sayood (1996) have explained a simple adaptive prediction scheme that exploits temporal correlations or spectral correlations in addition to spatial correlations. The hybrid scheme gives significant improvement in performance over other techniques. Besides prediction schemes, some simple error modeling techniques that take into account the prediction errors made in spectrally and/or temporally adjacent pixels is also considered. Efficient encoding scheme to encode the prediction residual is also described.

Chang-Hsing Lee and Ling-Hwei Chen (1997) have used the block sum pyramid to eliminate unnecessary search positions. It first constructs the sum pyramid structure of a block. Successive elimination is then performed hierarchically from the top level to the bottom level of the pyramid. Many search positions can be skipped from being considered as the best motion vector and, thus, the search complexity can be reduced.

A new fast algorithm for block-based motion estimation, the Flexible Triangle Search (FTS) algorithm, is presented in Mohamed Rehan et al (2007). The proposed algorithm is highly flexible due to its ability to quickly change its search direction and to move towards the target of the search criterion. It is also capable of increasing or decreasing its search step size to allow coarser or finer search. Unlike other fast search algorithms, the FTS can escape from inferior local minima and thus converge to better solutions.
Jo Yew Tham et al (1998) have proposed a novel Unrestricted Center-Biased Diamond Search (UCBDS) algorithm and it is more efficient, effective, and robust than the previous techniques like FSA, TSS NTSS and FSS. It has a best case scenario of only 13 search points and an average of 15.5 block matches. This makes UCBDS consistently faster than the other suboptimal block-matching techniques.

Two Cross-Diamond-Hexagonal Search (CDHS) algorithms, are proposed in Cheung and Po (2005) which differ from each other by their sizes of hexagonal search patterns. These algorithms basically employ two cross-shaped search patterns consecutively in the very beginning steps and then uses diamond-shaped patterns. To further reduce the checking points, two pairs of hexagonal search patterns are proposed in conjunction with candidates found located at diamond corners.

Po and Ma (1996) have explained a new Four-Step Search (4SS) algorithm with center-biased checking point pattern for fast block motion estimation. Halfway-stop technique is employed in the new algorithm with searching steps of 2 to 4 and the total number of checking points is varied from 17 to 27.

Lin and Gray (2004) have presented a new wavelet video coding algorithm and an optimization framework that allocates bit efficiency among the consecutive frames at the pixel level. The video residual coder described by them is based on the SPIHT and wavelet blocks allowing flexible bit allocation among active and inactive regions in a video frame. They have used Lagrange methods for optimizing the encoder for effective bit allocation.
Shamim and Robinson (2002) have described an object based video compression technique using detection and efficient coding of motion boundaries. In their technique, small numbers of global movement classes are first identified. Each of these classes is represented by two or more motion parameters. Then the regions of spatial segmentation are assigned to the motion classes and these segments are merged using similar heuristics. Finally, boundaries of motion are coded using an efficient asymmetrical binary coding scheme.

While the efficiency of exploiting interframe motion compensation is well established for lower bit rates, far less attention has been given to this issue for digital cinema resolutions. Michael Smith and John Villasenor (2004) address the specific coding efficiency differences between inter and intraframe coding for very high quality, high-resolution image sequences and a wide variety of bit rate ranges.

2.6 PATIENT VIDEO AND TELEMETRY

Liu et al (2004) have described a motion segmentation technique for efficient compression of patient video. They examine a Group Of Frames (GOF) to test whether a particular pixel was in the motion field or not. Insignificant pixel variations were ignored and threshold was set in order to detect the motion region.

Compressed video is necessary for a variety of telemetry requirements. A large number of competing video compression algorithms exist. Thom and Deutermann (2001) compare the ability of these algorithms to meet criteria which are of interest for telemetry applications. The algorithms are divided into those, which employ interframe compression, and those which employ intraframe compression.
Motion estimation is the most important part for exploiting the temporal redundancy between successive frames and it involves high computation cost. Yi et al. (2005) have proposed a hybrid coding, which uses the sum of absolute (transformed) differences (SA(T)D) as the block distortion measure and the proposed algorithm outperforms DS, 4SS, and NTSS search algorithms, having always better average computational complexity (slow, medium, fast motion) with similar performance to FS in terms of PSNR.

Belloulata and Shiping Zhu (2007) have proposed a novel object-based fractal monocular and stereo video compression scheme with quadtree-based motion and disparity compensation. Fractal coding is adopted and each object is encoded independently by a prior image segmentation alpha plane, which is defined exactly as in MPEG-4. The first n frames of right video sequence are encoded by using the Circular Prediction Mapping (CPM) and the remaining frames are encoded by using the Non Contractive Interframe Mapping (NCIM). The CPM and NCIM methods accomplish the motion estimation/compensation of right video sequence. According to the different coding or user requirements, the spatial correlations between the left and right frames can be explored by partial or full affine transformation quadtree-based disparity estimation/ compensation, or simply by applying CPM/NCIM on left video sequence.

Reducing the search complexity of matching between range block and domain block in fractal image compression is one of the most active research areas lately. Lifeng Xi and Liangbin Zhang (2007) have proposed an improved genetic algorithm for obtaining matching domain blocks to prevent the premature convergence of GA and to improve GA’s search efficiency. For each range block, an improved GA’s strategy is to encode the position of
searching domain block with Gray code, define fitness for minimum distance of current range block matching with searching domain block, select optimal individual protection strategy and adjust GA’s controlled parameters of crossover and mutation probability adaptively.

Shot detection and frame extraction techniques are illustrated in the following chapter.