Project 4

MPEG Video Encoder and Decoder

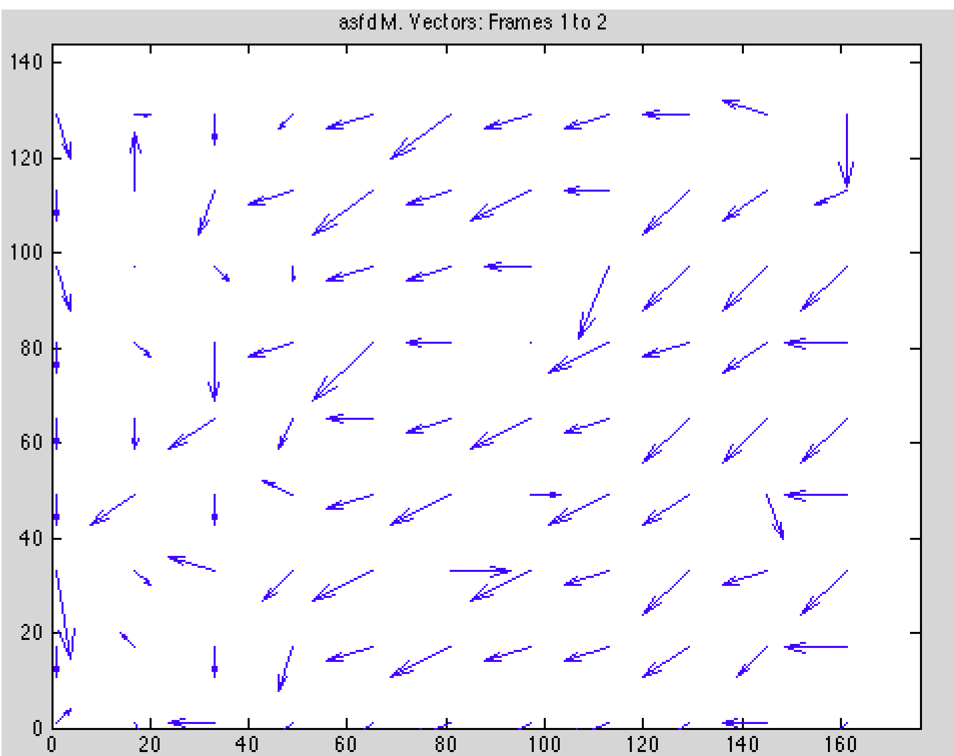
Alex Egg 809236396

COMPE565 Sprint 2014
Dr. Sunil Kumar

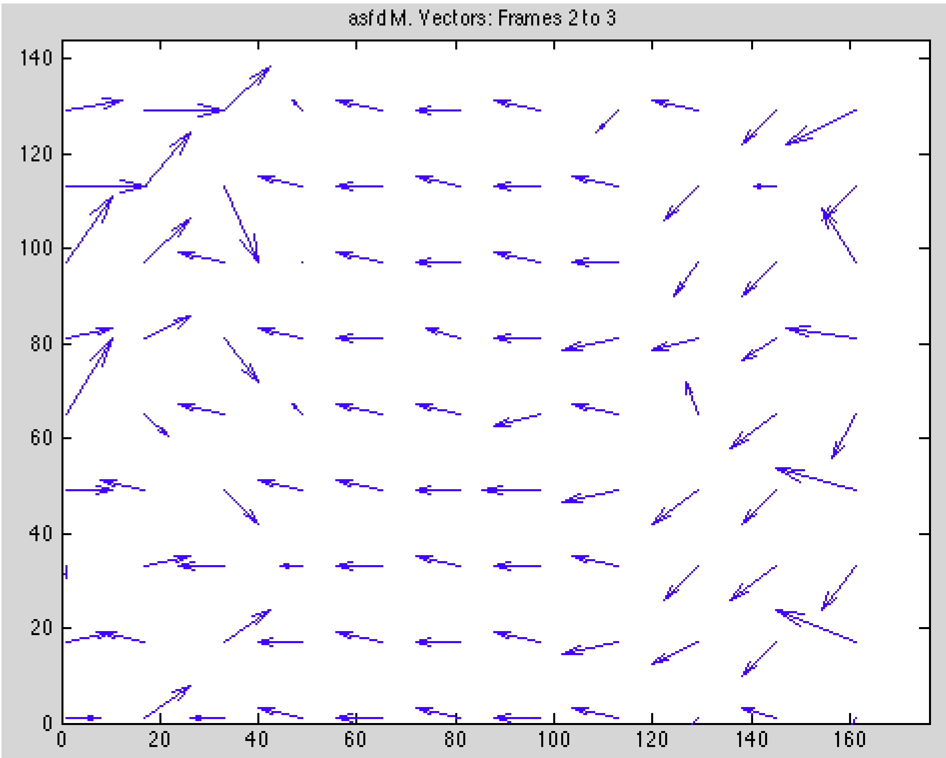
Table of Contents

List of Figures

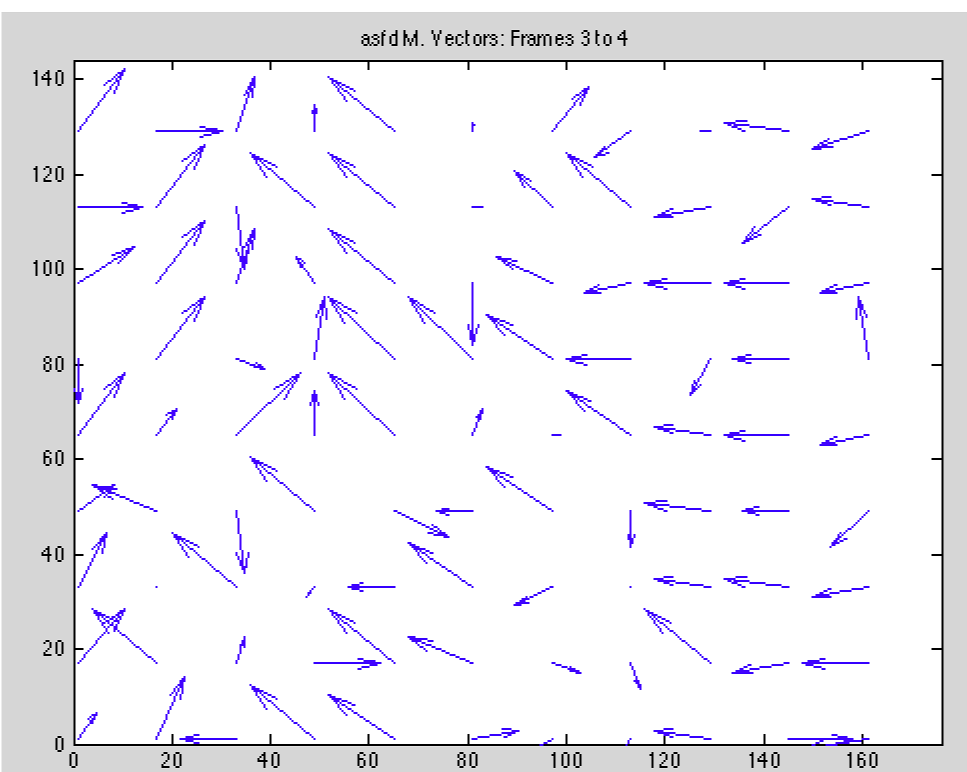
i. Motion Vectors for Frame 2



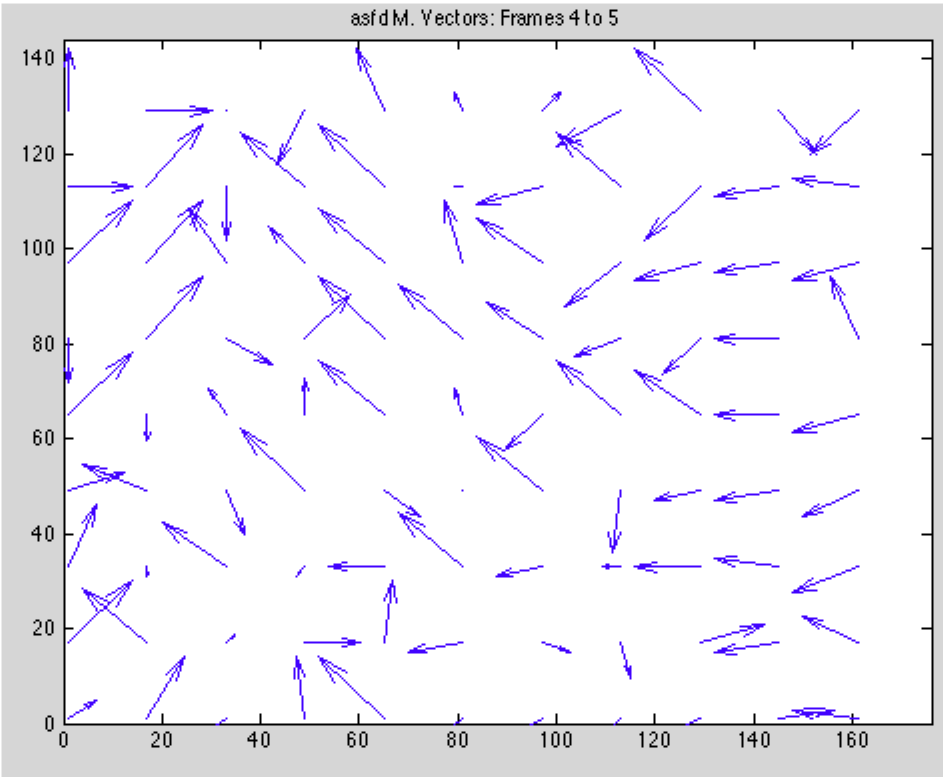
ii. Motion Vectors for Frame 3



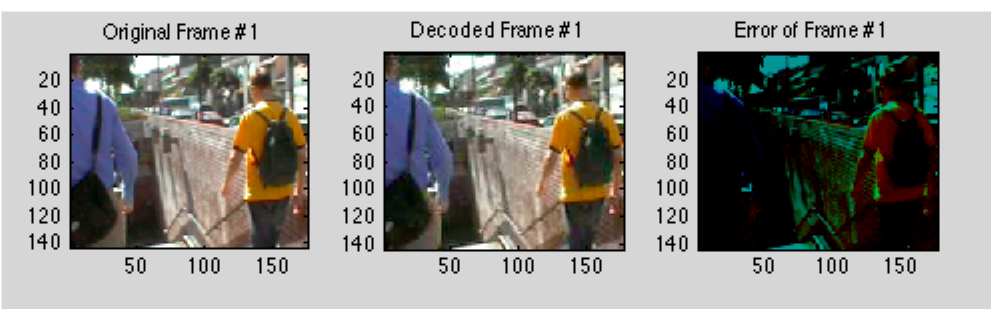
iii. Motion Vectors for Frame 4



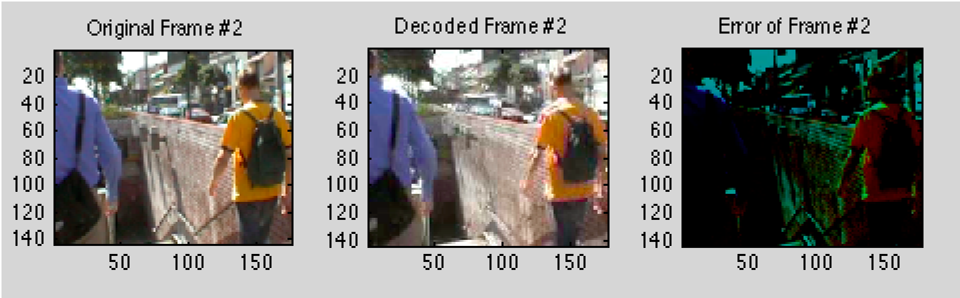
iv. Motion Vectors for Frame 5



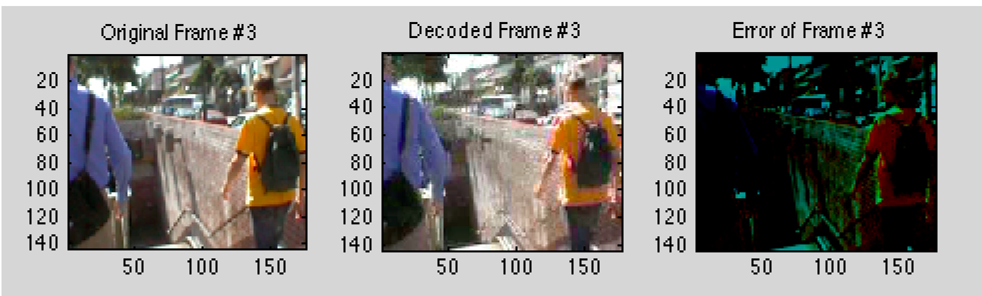
v. Original, Decoded, and Error for Frame 1



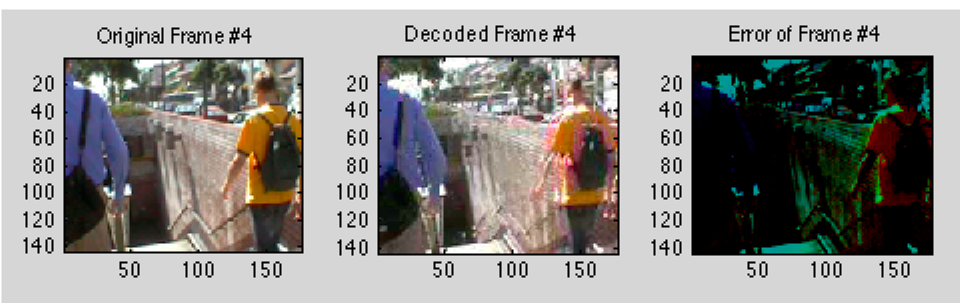
vi. Original, Decoded, and Error for Frame 2



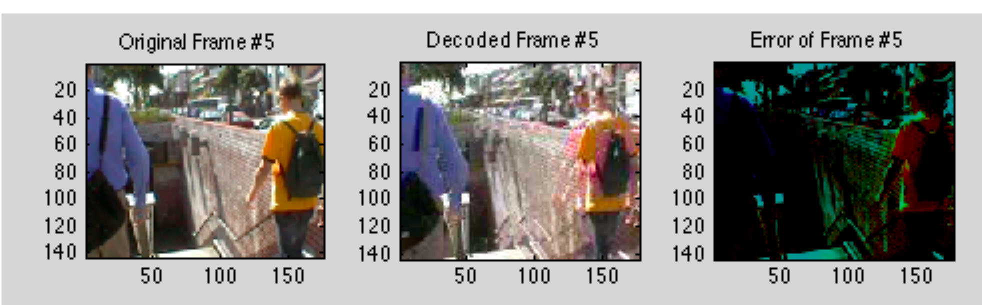
vii. Original, Decoded, and Error for Frame 3



viii. Original, Decoded, and Error for Frame 4



ix. Original, Decoded, and Error for Frame 5



Introduction

Video compression technologies are about reducing and removing redundant video data so that a digital video file can be effectively sent over a network and stored on computer disks. With efficient compression techniques, a significant reduction in file size can be achieved with little or no adverse effect on the visual quality. The video quality, however, can be affected if the file size is further lowered by raising the compression level for a given compression technique.

Different compression technologies, both proprietary and industry standards, are available. Most network video vendors today use standard compression techniques. Standards are important in ensuring compatibility and interoperability. They are particularly relevant to video compression since video may be used for different purposes and, in some video surveillance applications, needs to be viewable many years from the recording date. By deploying standards, end users are able to pick and choose from different vendors, rather than be tied to one supplier when designing a video surveillance system.

Axis uses three different video compression standards. They are Motion JPEG, MPEG-4 Part 2 (or simply referred to as MPEG-4) and H.264. H.264 is the latest and most efficient video compression standard.

We primarily consider lossy compression that yields perceptually equivalent but not identical video compared to the uncompressed source. There are 4 types of redundancies in such compression:

1. Perceptual
2. Temporal
3. Spatial
4. Statistical Redundancies

The MPEG standard defines 3 types of frames:

i) Intra Frame (I)

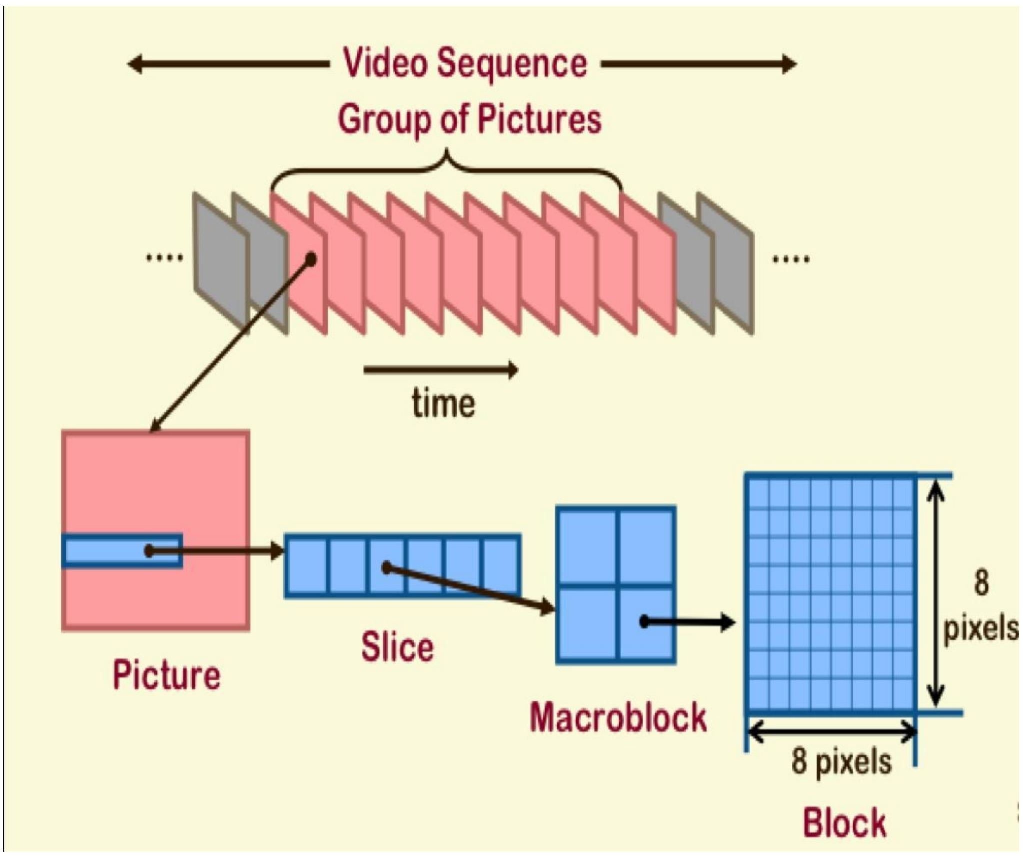
Refers to the various lossless and lossy compression techniques which are performed relative to the information that is contained only within the current frame and not relative to any other frame in the video sequence.

ii) Predicted Frame (P)

Refers to images within a sequence of images that are created using information from other images from I frames. They also “forward” predict frames. However you must have decoded the previous frame in order to make any use of a P frame.

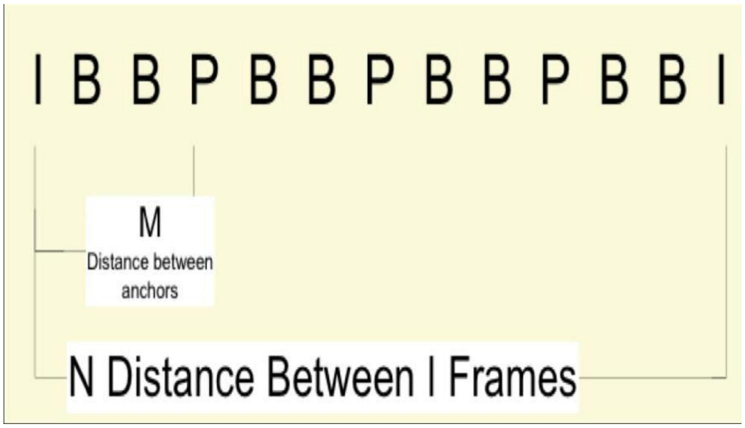
iii) Bidirectional Frame (B)

Refers to frames which store the difference between past and future frames. The most compatible frame type, holding least information typically used in fast action sequences. This is the most compressible frame type because it holds the least amount of information, similar to P frames.



MPEG Video Stream Structure

Group Of Pictures (GOP): The distance between I frames is known as a group of pictures' size. A GOP consists of all the pictures that follow a GOP header before another GOP header. The GOP layer allows random access because the first picture after the GOP header is an Intra picture which means that it doesn't need any reference to any other picture.



Macro Block

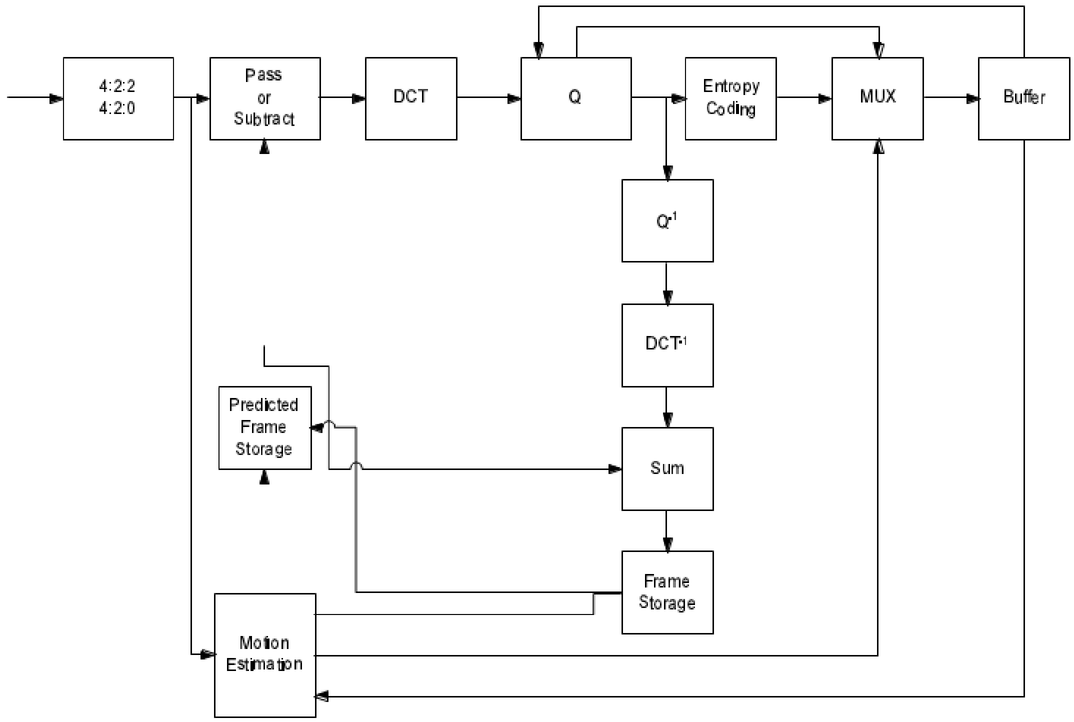
Macroblock: Basic coding unit in the MPEG algorithm. The size is usually 16x16 but other block sizes can also be utilized. The size of a Macroblock is important in accuracy and speed of the motion estimation operation. The smaller the macroblock size, the higher the side information generated and vice-versa.

Block: Smallest coding unit in the MPEG algorithm. Consisting of 8x8 pixels and can be one of the 3 types: luminance (Y), red chrominance (Cr), or Bleu Chrominance (Cb). The block is the basic unit in intra frame coding. The DCT or Discrete Cosine Transform is performed at the block layer.

Objective

Develop a basic MPEG Video encoder and Decoder to understand the working principles of video coding.

Procedural Section



MPEG Encoder Architecture

The steps involved for the Decoder is as follows:

i) 4:2:0 Sub Sample

ii) Motion Estimation

iii) Exhaustive Search

iv) DCT

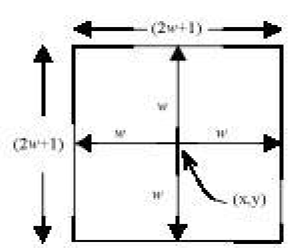
v) Quantization

4:2:0 Sub Sample: The human eye is more sensitive to luminance than color. We can exploit this natural fact by transmitting more luminance and reducing the detail in chrominance to accommodate storage and bandwidth limitations. In the 4:2:0 sub sampling scheme YCbCr consists of 2 luminance samples for every Cb and Cr sample but there is no chrominance sample in the alternative rows. Overall, there are no chrominance samples in the alternative rows and columns.

Motion Estimation: There are many ways Motion estimation can be achieved. We used exhaustive search and three step search method in this project to achieve our goals. Motion estimation over a macroblock of pixels is a standard approach for estimating motion in a moving image sequence (video). There are 4 important parameters:

1. Macroblock size
2. Matching Criteria
3. Prediction Mode
4. Search Technique

Exhaustive Search (ES) : In Exhaustive Search (ES), every possible displacement within a rectangular search window is attempted. The displacement that produces the minimum distortion is chosen as the motion vector. As shown in Figure 1, ff the maximum search range in either direction is w (assuming a square search).



Three Step Search: Like most algorithms in live there is a naive approach and usually more sophisticated approaches. If we just add a heuristic to our algorithm we can achieve a faster runtime and lower computational complexity. There are various algorithms that improve upon exhaustive search, but the one I chose to implement is TSS. This is one of the earliest attempts at fast block matching algorithms and dates back to mid 1980s. The general idea is represented in Figure 2. It starts with the search location at the center and sets the ‘step size’ S = 4, for a usual search parameter value of 7. It then searches at eight locations +/- S pixels around location (0,0). From these nine locations searched so far it picks the one giving least cost and makes it the new search origin. It then sets the new step size S = S/2, and repeats similar search for two more iterations until S = 1. At that point it finds the location with the least cost function and the macro block at that location is the best match. The calculated motion vector is then saved for transmission.

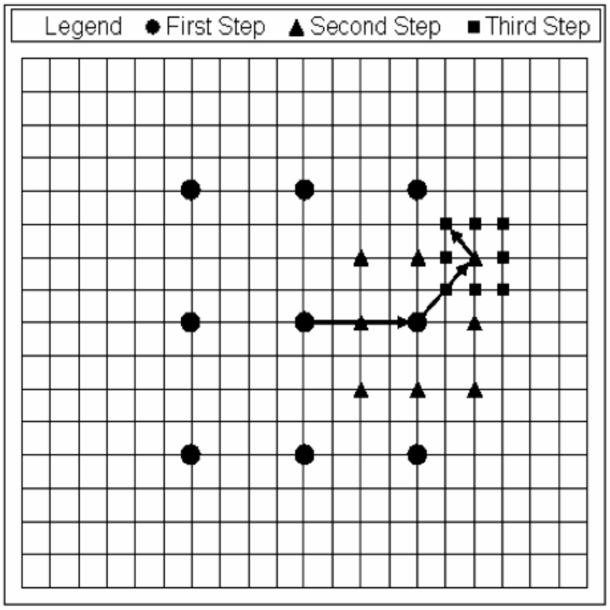


Figure 2: Three Step Search procedure. The motion vector is (5, -3).

Quantization: In quantization the low frequency components are scaled down by a small value while the higher frequency components are scaled down by a larger factor. The DCT transformed coefficients for Y, Cb, and Cr have most energy at the low frequency i.e. upper Left corner of 8x8 block and low energy at higher frequency i.e. lower right values. The low energy Coefficients are often small enough to be neglected with little visible distortion. This is done through quantization.

The steps involved for the Decoder is as follows:

i) Inverse Quantization

ii) Inverse DCT

iii) Motion Compensation

iv) Up sampling

Inverse Quantization: It is the first step in the decoder to reconstruct the image.

Inverse DCT: It is the reverse process of DCT. The original frame and the reconstructed frame will be identical without quantization but might differ a little due to rounding errors.

Motion Compensation: Each of the frames is reconstructed at the decoder with the help of the reference frame.

Up sampling: The YCbCr image components are recovered after the IDCT is performed. Then the image is displayed in the RGB format by Up-sampling the image from 4:2:0 to 4:4:4.

Conclusion

The MPEG encoder and decoder are developed and implemented. In this we exploit different redundancies and achieve maximum compression ratio. As MPEG is a lossy compression technique, we observe blocky nature in the reconstructed frames at the decoder side. We saw the difference between the original frame and reconstructed frame with the difference frame.

References

1. http://www.axis.com/products/video/about_networkvideo/compression.htm
2. Dr. Sunil Kumar Lectures
3. Wikipedia