CHAPTER 3

METHODOLOGY

From the massive development in the digital communication medium, it is obvious that multimedia contents like images, video and audio, needs a method capable of offering a good content protection. The present research work proposes watermarking methodology based on the effective, synergistic integration of several schemes that

- efficiently selects optimal frames that produce minimum distortion after embedding the copyright information.
- embed the copyright information in an uncompressed raw video.
- embed the copyright information in an compressed MPEG video.

New techniques are proposed because of the following reasons :

- Researchers should always challenge industrial standards by developing new competitive methodologies.
- Moving towards digital offices, otherwise several business and industrial applications require secure watermarked videos that can be safely used and transmitted.

In order to meet the objectives formulated in Chapter 1, the research design (Figure 3.1) framed answers the following research questions.

- Where should the watermark be embedded ?
- How should the watermark be embedded ?
- How to detect and extract the watermark ?

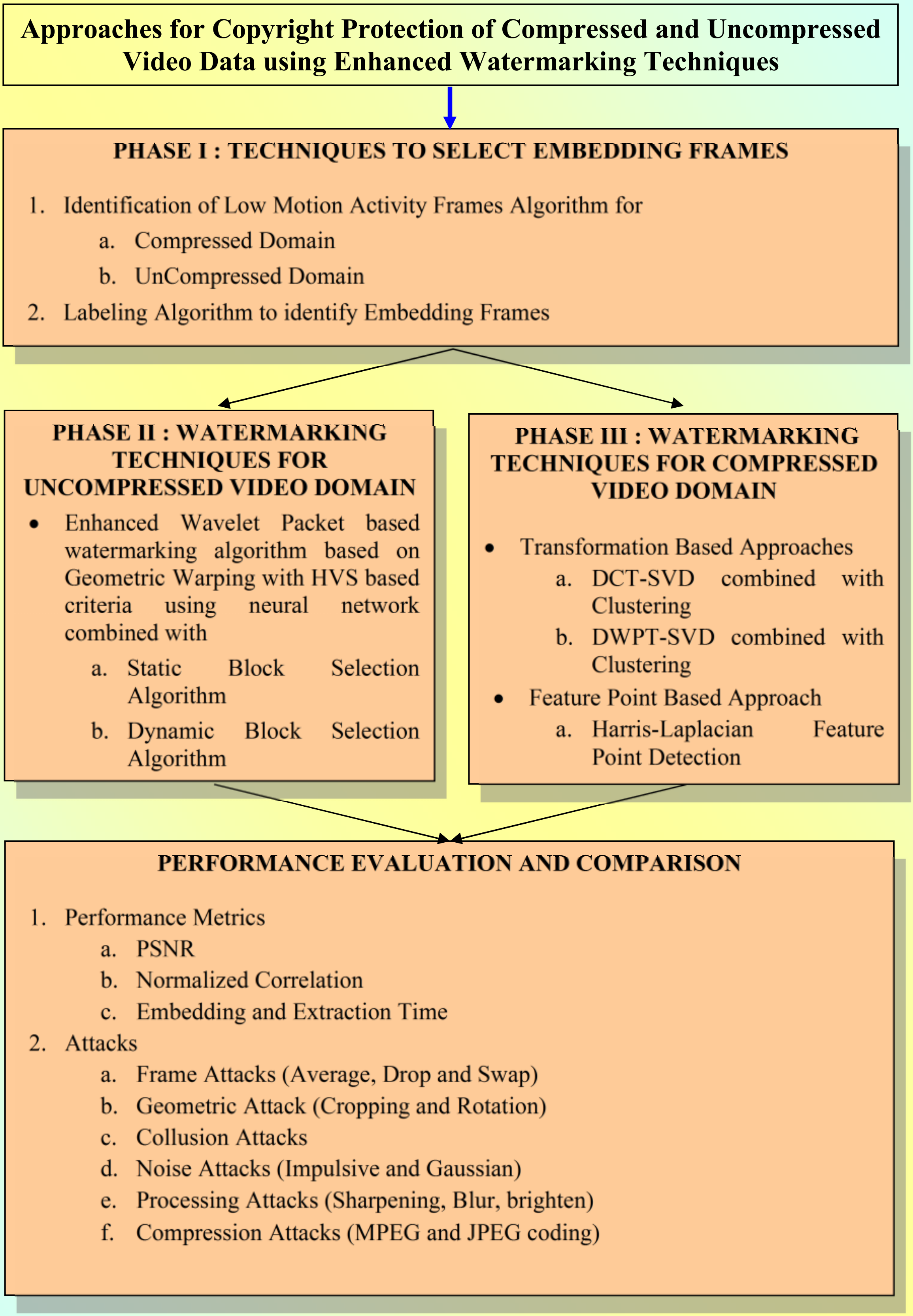

Figure 3.1 : Research Design

Phase I : Selects the place of embedding the copyright watermark

Phase II : Design and implement embedding and extraction algorithms for uncompressed video data

Phase III : Design and implement embedding and extraction algorithms for compressed video data

In order to increase the protection of copyright information, the embedding procedure uses a nested watermark, created using Visual Cryptography (VC) technique (Naor and Shamir, 1995). VC is a Secret Sharing Scheme (SSS) that has the ability to restore the secret data without the use of complex computations. Nested watermarking is the procedure of embedding a watermark into another watermark. In this research, the copyright image is divided into two shares using the basic (2, 2)-threshold VC scheme and the second share is embedded into first share, thus creating a new nested watermark having the copyright information. This image is then embedded into the compressed or uncompressed video data.

3.1. TECHNIQUES TO SELECT EMBEDDING FRAMES

Phase I of the research work focuses on the selection of regions for inserting the watermark. The objective of Phase I is to find the frames where insertion of watermark has minimum distortion and maintains transparency. The procedure consists of two steps. The first step identifies low motion activity frames where the insertion of watermark will not be visible to human visual system and the second algorithm selects frames from this subset that has minimum impact on visual quality after insertion of a watermark.

Two separate algorithms, termed as ILMAF-C (Identification of Low Motion Activity Frames Algorithm for compressed domain) and ILMAF-UC (Identification of Low Motion Activity Frames Algorithm for Uncompressed domain), are proposed for this purpose. To identify low activity frames, a motion activity measure, that describes the level of intensity of activity, action or motion in the video sequence, is used. This measure is estimated from the motion intensity and motion activity matrix stored with the MPEG compressed video. In the uncompressed domain, this measure is calculated using a block matching method that uses an enhanced adaptive rood pattern search algorithm to estimate motion intensity. After estimating the motion activity, the frames are categorized as very low, low, medium, high and very high activity frames. Only the very low, low and medium frames are selected for embedding process.

A labeling algorithm is then used to identify the embedding regions in each of these frames. The algorithm treats the selected frames as a sequence of images and segments each frame as Embed Region (ER) and Other Region (OR). The ER is further divided into three regions namely, low complex (smooth), medium complex and high complex regions. By determining the amount of smooth, medium and high complex regions in the frames, a frame is respectively identified as smooth, medium or high complex frames. During embedding, no data is embedded in the high complex frame and 2/3rd of embedding is done in smooth and rest is done in medium frames.

Experimental results proved that the insertion of phase I techniques improved security, transparency and robustness of the proposed watermarking algorithms.

3.2. WATERMARKING TECHNIQUES FOR UNCOMPRESSED VIDEO DOMAIN

The techniques proposed in Phase II of the research work are designed and developed for inserting watermarks in uncompressed video data using wavelet packet transformation. While watermarking uncompressed video data, four issues are identified with the existing solutions.

(i) The embedding procedures used are non-block based schemes where the same technique is used in all frames, which reduces the resistance against attacks.

(ii) Watermarks are generally embedded into the perceptual invisible part and redundant data of the video. Both video watermarking and compression algorithms focus on this region, where watermark algorithm attempts to embed data while compression algorithm removes them to reduce the amount of storage space used to store the video. Thus, contradiction arises.

(iii) Irrespective of enhanced techniques used, existing algorithms introduce artifacts and distortions that degrade the visual quality of the video data.

(iv) Almost all the algorithms require the original video data to be present to extract the copyright information

The research work proposes enhanced Wavelet packet based watermarking algorithms based on Geometric warping with HVS (Human Visual System) based criteria that use neural network combined with Static Block Selection Algorithm (WGSBS) and Dynamic Block Selection Algorithm (WGDBS). The algorithm using the frames selected in Phase I, first performs DWPT to obtain subbands, which are first segmented into non-overlapping 8 x 8 blocks. Just Noticeable Difference (JND) (Damnjanovic and Izquierdo, 2006) HVS characteristics are then calculated to obtain the allowable visibility ranges for all coefficients of a wavelet-packet transformed image. The inclusion of JND increases the imperceptibility of the algorithm. An Artificial Neural Network (ANN) is then used to memorize the relationships between the original wavelet coefficients and its watermark version. A geometric warping method (Pröfrock *et al.*, 2007) then embeds the watermark in the selected regions of the frames. An inverse DWPT produces the watermarked image.

The geometric warping method is enhanced through the use of a block selection algorithm. The block selection algorithm aims to reduce the artifacts or flickering effects introduced by the original warping method. The study proposes a static and dynamic block selection algorithm for this purpose. Static block-based watermarking techniques consist of dividing the image into non-overlapping blocks of pixels and inserting a watermark into each block. A spatial and feature based clustering method is used to create the dynamic blocks. The method is devised to group pixel data by taking into account simultaneously both their feature space similarity and spatial coherence.

The extraction algorithm, similar to embedding algorithm, first selects the frames from the watermarked video and performs static or dynamic block segmentation. DWPT is performed and these coefficients along with the trained ANN are then used to compute and extract the nested watermark. From the extracted watermark, first share two is removed from share one and then both the shares are XORed to obtain the original copyright image.

3.3. WATERMARKING TECHNIQUES FOR COMPRESSED VIDEO DOMAIN

The third phase of the study proposes two transformation techniques and one feature point based technique for watermarking compressed video data.

3.3.1. Transformation Based Approaches

In the first category, the embedding procedure combines Discrete Cosine Transformation (DCT) or Discrete Wavelet Packet Transformation (DWPT), with Singular Value Decomposition (SVD). The dynamic block selection algorithm used in Phase II algorithms are again used here to select the region of embedding in each frame obtained from Phase II.

The algorithm first performs a DCT or DWPT transformation obtains coefficients, from which the SVD values are obtained. The watermark is embedded in the U component of SVD values to increase robustness. Finally inverse SVD followed by IDCT or IDWPT is performed to obtain the watermarked image. The extraction process follows the same steps to extract the nested watermark, from which using VC concept the copyright image is obtained.

3.3.2. Feature Point-Based Algorithm

This algorithm, after selecting frames for embedding (Phase I), performs quad-tree decomposition on each frame to obtain representations at different resolution levels. In the next step, feature points extraction and Local Circular Regions (LCRs) construction are performed using Harris-Laplace detector (Deng *et al.*, 2010). A mechanism of minimum distance clustering-based feature selection is used to select a set of non-overlapped LCRs from which geometrically invariant LCRs are completely formed. The histogram and the mean statistically independent of the pixel positions are calculated over the selected LCRs, which are then utilized to embed the watermark.

The embedding quality is improved by identifying bad bins using a technique that analyzes the relative relationship between groups of two neighboring bins. The number of pixels in the bin has essential effect on the relative relationship. If this number is equal to zero or when it is less than a threshold value, then the concerned bins are considered as bad bins. The algorithm then removes these bad bins in LCR that has a negative impact on embedding quality. The extraction procedure uses a similar procedure to extract the nested watermark from the watermarked video, from which the copyright image is extracted using visual cryptography.

3.4. TECHNIQUES USED

The study uses two main transformation techniques, namely, Discrete Wavelet Packet Transformation (DWPT) and Discrete Cosine Transformation (DCT) along with visual cryptography during the design of the proposed video watermarking techniques. This section presents a description on the working of these techniques.

3.4.1. Discrete Cosine Transformation

Discrete Cosine Transformation (DCT) translates the image information from spatial domain to frequency domain to be represented in a more compact form. Its stochastic properties are similar to Fourier transformation and consider the input image to be a time invariant or stationary signal. The DCT is a special case of Discrete Fourier transformation (DFT) in which the sine components have been eliminated leaving only the cosine terms (Mandyam *et al.*, 1997).

The forward N-point one-dimensional DCT and inverse DCT can be defined as in Equations (3.1) and (3.2).

$$\text{Forward N-point DCT} = X(k) = \frac{2}{N} c_k \sum_{n=0}^{N-1} x(n) \cos\left[\frac{(2n+1)k\pi}{2N}\right], \quad (3.1)$$

k=0,1,………,N-1

Where N is the size of the DCT block, X (with x data signals) is the set of data signals inside the DCT Kth block, denotes the kth component and k=0,1,………,N-1, where n is the number of signals.

$$\text{Inverse N-point DCT} = x(n) = \frac{2}{N}\sum_{k=0}^{N-1} c_k X(k)\cos\left[\frac{(2n+1)k\pi}{2N}\right], \quad (3.2)$$

n=0,1,………,N-1

where $c_k = \begin{cases} 1/\sqrt{2}, k = 0 \\ 1, k \neq 0 \end{cases}$.

Both DCT and IDCT are orthogonal, separable and real transforms. Being separable means that the multidimensional transformation can be decomposed into successive application of one-dimensional transforms in the appropriate directions. Similarly orthogonal means if the matrices of DCT and IDCT are non-singular and real then their inverse is obtained merely by applying transpose operation. Like Fourier transformation, DCT also considers the input sampled data to be a time invariant or stationary signal.

In case of image and video compression standards such as baseline JPEG and MPEG, 8-point DCT and IDCT are used. The image or each motion video frame of size NXN pixels is divided into two-dimensional non-overlapping blocks often called sub-images or basis functions of size 8x8 (having 64 pixels each) and 2-D DCT is applied on the encoder side, while on the decoding side 2-D IDCT (2-dimensional Inverse Discrete Transformation) is applied to recover the original data.

The 8-point 2-D DCT and IDCT to generate 8x8 data matrices are calculated as:

$$\text{2-D DCT} = X_{k,l} = \frac{c(k)c(l)}{4}\sum_{m=0}^{7}\sum_{n=0}^{7} x_{m,n}\cos\left(\frac{(2m+1)k\pi}{16}\right)\cos\left(\frac{(2n+1)l\pi}{16}\right), \quad (3.3)$$

where k,l = 0,1,………..,7

$$\text{2-D IDCT} = x_{m,n} = \sum_{k=0}^{7}\sum_{l=0}^{7}\frac{c(k)c(l)}{4} X_{k,l}\cos\left(\frac{(2m+1)k\pi}{16}\right)\cos\left(\frac{(2n+1)l\pi}{16}\right) \quad (3.4)$$

where m,n = 0,1,………..,7 and

$$c(k),\ c(l) = \begin{cases} 1/\sqrt{2}, k \ \&\ l = 0 \\ 1, \text{otherwise} \end{cases} \qquad (3.5)$$

One strategy to compute the 2-D DCT and IDCT is the standard row-column separation. The 2-D transformation is performed by applying the 1-D transformation to each row and subsequently to each column of the data matrix.

3.4.2. Discrete Wavelet Packet Transformation

The DWPT can be regarded as any one of a collection of orthonormal transforms, each of which can be readily computed using a very simple modification of the pyramid algorithm for the DWT.

Since images are typically a composition of mid and high spatial frequency components, the wavelet decomposition does not capture its typical structure. These structures are better found by wavelet packets. The wavelet packet technique was introduced in (Coifman and Wickerhauser, 1992**)** as a natural extension of the wavelet techniques and was immediately investigated in various areas of signal analysis.

The wavelet packet basis is constructed adaptively based on cost functions and choice of decomposition filter applied to an image. The decomposition of a signal can be viewed as a tree, where the left branch represents the low-pass horizontal / low pass vertical filtering, the right branch represents the high pass horizontal / high pass vertical filtering and the middle branches represent the low-pass horizontal/high pass vertical filtering and the high pass horizontal/low pass vertical filtering respectively.

The wavelet packet tree is constructed in two steps. First, the image is filtered and subsampled into four new images representing the spatial frequency subbands. Each of the subbands are filtered and subsampled into four new images, a process that is repeated to a certain level. By keeping the components in every subband at every level the wavelet packet tree has the advantage of attaining the complete hierarchy of segmentation in frequency and is a redundant expansion of the image. This is illustrated by the subband stack in Figure 3.2 which shows the subbands of different scales, the chosen subbands in grey form the resulting orthogonal basis shown at the bottom, to the left a wavelet basis and to the right a wavelet packet basis.

In the second step, the best basis is chosen to represent the image by cutting off branches in the tree controlled by the cost function applied on one node and on its children nodes. These subbands are to be seen as a collection of bases for this particular image and filter choice. Choosing a best basis is a way to represent the image in the most effective way.

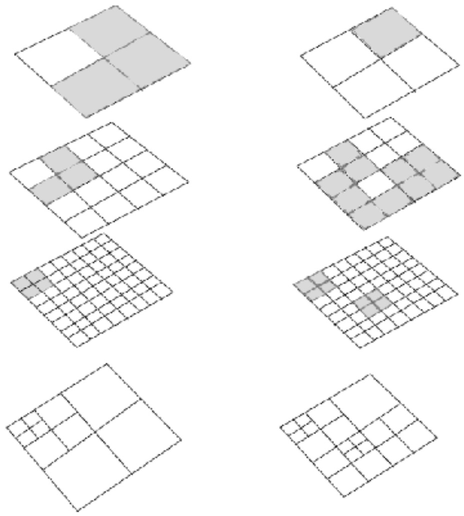

Figure 3.2: The Subband Stack

The fully decomposed wavelet packet tree is shown in Figure 3.3. The wavelet packet transformation generates an orthogonal basis if a proper wavelet is chosen.

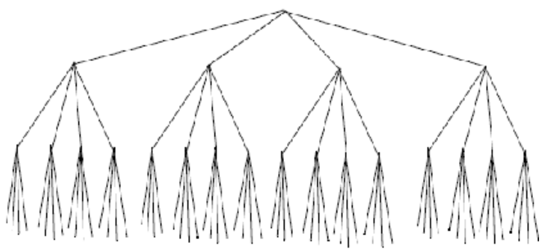

Figure 3.3 : Fully Decomposed 3-level Wavelet Packet Tree

For basis selection, a cost function M(A) is applied on the unsplit subband A and the same cost function is also applied to all subbands that result from one level subband decomposition. This process is illustrated in Figure 3.4. The subband is split if M(A) > M(B) + M(C) + M(D) + M(E), otherwise the subband is kept as it is.

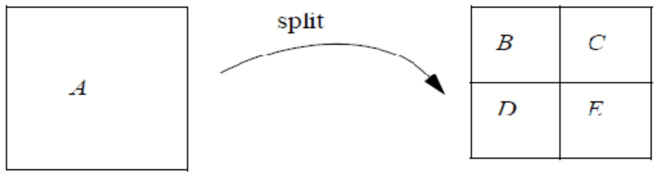

Figure 3.4 : One Step Subband Split

Figure 3.5 shows two pruned trees, the left representing the wavelet basis and the right representing a possible best basis. The leaves of the tree represent the basis. Wavelet packet decomposition of an image and basis selection results in different bases when using different filters or different cost functions. The best basis found for a specific image, using a particular filter and cost function, is best for this triplet of image, filter and cost function. Changing atleast one of the three may result in another best basis.

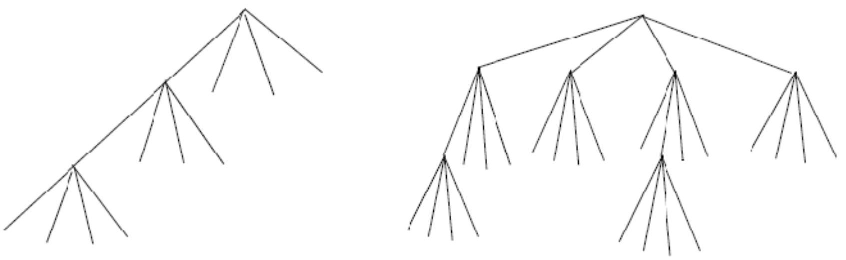

Figure 3.5 : Tree Representation of the Bases

3.4.3. Visual Cryptography

Visual Cryptography (VC) is a Visual Secret Sharing Scheme (VSSS) introduced by Naor and Shamir (1995) for encrypting materials like written text, printed text, pictures, in a secure fashion. It uses the Human Visual System (HVS) to decrypt a secret message without expensive and complicated decoding process (Tai and Chang, 2005).. A VSSS partitions an image encoded with a secret digital signal 'S' into a collection containing 'n' black and 'm' white pixels. Each collection of m x n pixels is referred to as a share, which resembles a noisy and scrambled image when viewed separately. During decoding phase, these shares or subset of shares are stacked together to allow the visual recovery of the secret message.

VC has been applied to many applications, including but not restricted to E-voting system (Paul *et al.*, 2003), financial documents (Hawkes *et al.*, 2000), information hiding (Bonnis and Santis, 2004), general access structures (Ateniese *et al.*, 1996), visual authentication and identification (Naor and Pinkas, 1997). More detailed information about visual cryptography can be found in (Yang, 2002). The solutions normally operate on binary or binarized inputs. VC technique has many advantages as listed below.

- Simple to implement and low cost algorithm
- Encryption does not require any heavy computation
- No separate decryption algorithm is required (uses only Human Visual System) and thus, no prior knowledge of cryptography is needed during decrypt.
- Any share by itself or using a subset of shares does not provide enough information but together, overlaid in correct order reveals S.
- Highly secure method, even with the help of infinite computation power, it is very difficult to predict S.

Visual cryptographic solutions normally operate on binary or binarized inputs. After its initial introduction, many researchers have found different variations of VC (Yang, 2010). The improvement varies from binary image to grey scale and colour images. In halftone VC, the natural (continuous-tone) images are first converted into halftone images by using the density of the net dots to simulate the original grey or color levels in the target binary representation. Then, the halftone version of the input image is used instead of the original secret image to produce the shares. A half tone image is the binary version of the grey scale image. The halftoning technique is used in many applications such as facsimile (FAX), electronic scanning and copying, and laser printing etc. The decrypted image is obtained by stacking the shares together.

Visual Cryptographic systems are being used in various applications such as E-voting system, financial documents and secure image transmission. In recent years, it is also used with another important technique ‘Watermarking’. VC used in conjunction with watermarking allows multiple watermarks to be embedded in the same image without modifying the host image (Luo *et al.*, 2008; 2009). In addition, it has the advantage that the watermarks can be extracted without using the original image. Thus, they are very suitable for many applications including medical images and financial document images.

This research work uses the basic (2, 2)-threshold VSSS. The procedure (Figure 3.6) is explained below.

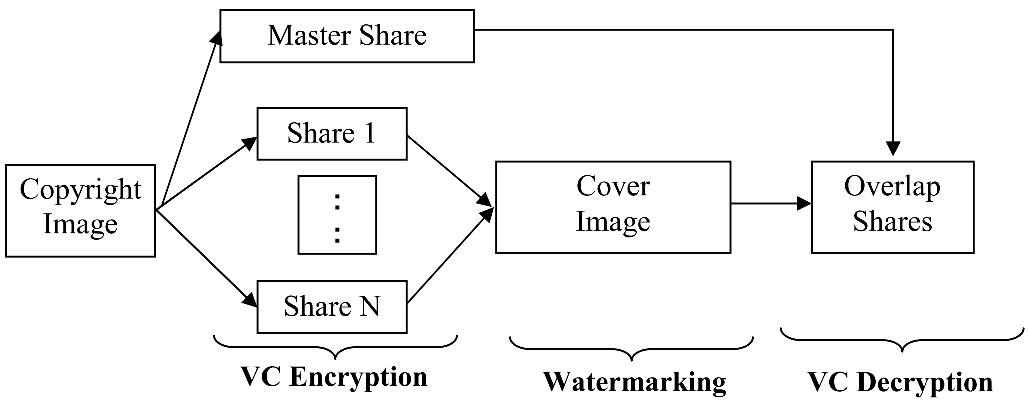

Figure 3.6 : Visual Cryptography Procedure

In this algorithm, each pixel of the binary secret image is expanded into 2 * 2 pixels (Figure 3.7). To share a white pixel of the secret image, one row from the first 6 rows of Figure 3.7 is chosen randomly. Similarly, the two shares of a black pixel are determined by a random selection from the 6 last rows of Figure 3.7. As a result, an M*N pixels secret image is expanded into two 2M*2N pixels share-images.

Considering security of the method, the presence of only one share image reveals nothing about the corresponding secret image, i.e., each 2*2 pixels block of one share-image may correspond to either a white pixel or a black pixel of the secret image. Stacking the shares of a black secret pixel results in 4 black subpixels, whereas only 2 black subpixels is gained by stacking shares of a white secret pixel (Houmansadr and Ghaemmaghami, 2005). So, secret image is revealed to human eyes by stacking the shares without performing any cryptographical computations.

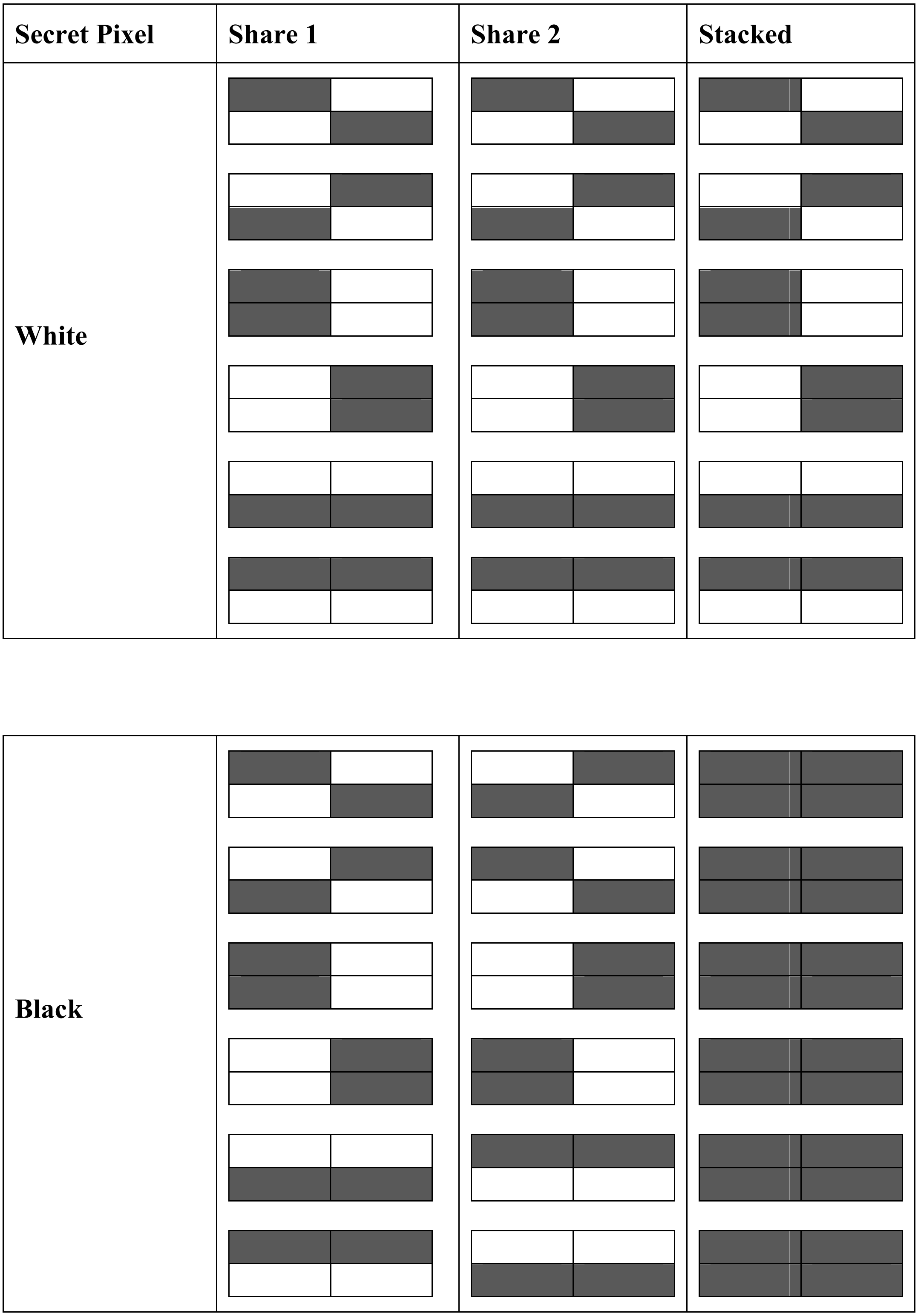

Figure 3.7 : (2, 2) VC Scheme with 2x2 Subpixels

One drawback of using (2, 2) VCSS is complete recovery of white pixels is difficult and thus has loss in contrast in the reconstructed image. This is due to the OR operation used during reconstruction. Therefore, a XOR based VCS scheme is generally used, where the images are superimposed using XOR operation. This results in perfect reconstruction of both black and white pixels as shown in Figure 3.8.

Secret Image	**Shares**	**OR**	**XOR**
White pixel (0)	$\begin{bmatrix}1 & 0\\1 & 0\end{bmatrix} \bullet \begin{bmatrix}1 & 0\\1 & 0\end{bmatrix}$	$\begin{bmatrix}1 & 0\\1 & 0\end{bmatrix}$	$\begin{bmatrix}0 & 0\\0 & 0\end{bmatrix}$
Black pixel (1)	$\begin{bmatrix}1 & 0\\1 & 0\end{bmatrix} \bullet \begin{bmatrix}0 & 1\\0 & 1\end{bmatrix}$	$\begin{bmatrix}1 & 1\\1 & 1\end{bmatrix}$	$\begin{bmatrix}1 & 1\\1 & 1\end{bmatrix}$

Figure 3.8 : Reconstruction of Pixels

3.5. MPEG VIDEO FORMAT

MPEG (Moving Picture Expert Group) video format is a de-factor compression algorithm used by many current and emerging video processing applications like digital television set-top boxes, HDTV decoders, video conferencing, DVD players and Internet video. It was introduced during 1988 by International Standards Organisation (ISO) to standardize video compression. This format has the advantage of producing a compressed video file that requires less storage and does not reduce quality. As MPEG video standard is used during the design of compressed watermarking domain, this section presents a description (Gilvarry, 1999) of this format.

3.5.1. MPEG-1 Bit Stream

The bit stream is in a layered format as shown in Figure 3.9, a brief description of the function of each layer is given in Table 3.1 and a brief description of the same is provided below.

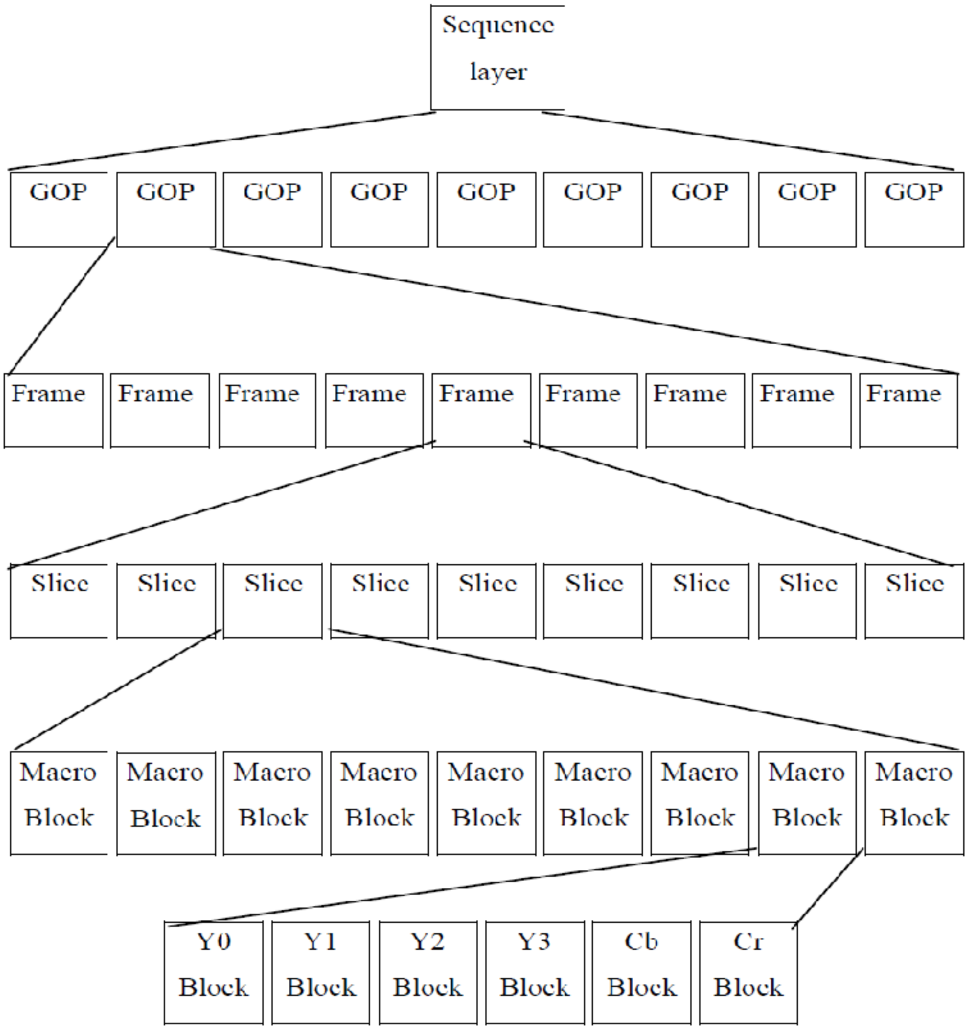

Figure 3.9 : Layered Structure of MPEG Bit Stream

TABLE 3.1

FUNCTION OF EACH LAYER OF THE BIT STREAM

Layer	**Function**
Sequence Layer	One or more groups of pictures
Group of Pictures (GOP)	Random access into the sequence
Picture	Primary coding unit
Slice	Resynchronization unit
Macroblock	Motion compensation unit
Block	DCT unit

1. The Sequence layer contains general information about the video, which includes the vertical and horizontal size of the frames, height/width ratio, picture rate, Buffer size, Intra and non-intra quantizer default tables.

2. Group of Pictures (GOP) layer groups together the pictures to support greater flexibility and efficiency in the encoder/decoder.

3. The Frame layer (picture layer) is the primary coding unit, it contains information regarding the picture's position in the display order (pictures do not come in the same order as they are displayed), what type of picture it is (Intra, Predicted or Bi-directionally predicted) and the precision and range of any motion vectors present in the frame.

4. The Slice layer is important in the handling of errors. If the decoder comes across a corrupted slice, it skips it and goes straight to the start of the next slice.

5. The Macroblock layer is the basic coding unit and within this unit the motion vectors are stored. Each macroblock may have one or motion vectors associated with it.

6. The Block layer is the smallest coding unit and it contains information on the coefficients of the pixels.

- **Description of a frame**

As mentioned above, there are three types of picture/frame:

- Intra (I-type) : These frames are encoded using only information from itself.

- Predicted (P-type) : These frames are encoded using a past I or P frame as a reference (Figure 3.10). This is known as forward prediction.

- Bi-directionally predicted (B-type) : These frames are encoded using a past (forward predicted) and a future (backward predicted) I or P frame as a reference (Figure 3.11). A B-type frame is never used as a reference.

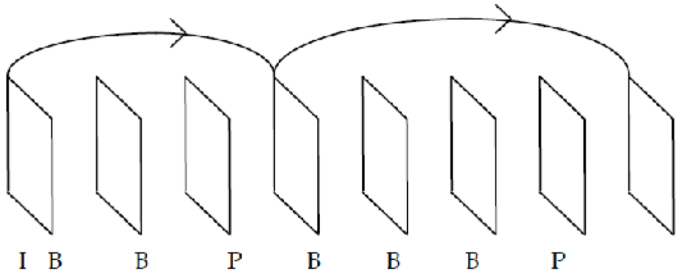

Figure 3.10 : P Frames

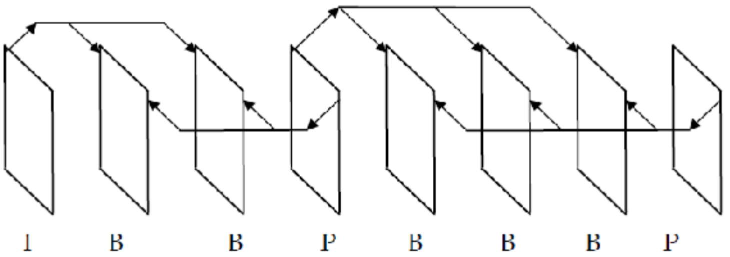

Figure 3.11 : B Frames

Each frame is divided up into arbitrary sized slices. A slice may contain just one macroblock or all the macroblocks in the frame. As shown in Figure 3.12, a slice is not confined to a single row.

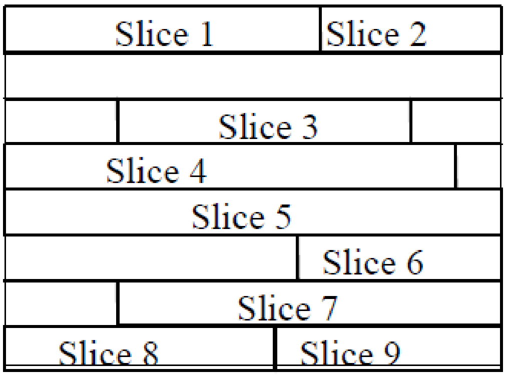

Figure 3.12 : A Single Frame Divided into Slices

- **Bit stream order and display order of frames**

A typical sequence of frames in the display order is shown below.

I	B	B	B	P	B	B	B	B	P	I	B	B	B	I
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

However, this is not the order in which they are transmitted! The P frame numbered five is needed for the decoding of B frames two, three, and four. Therefore five has to be decoded before two, three and four and hence transmitted before them. Similarly P9 is transmitted before B6, B7, B8 and B9 also I15 is transmitted before B12, B13 and B14. The bit stream order is shown below.

I	P	B	B	B	P	B	B	B	B	I	I	B	B	B
1	5	2	3	4	10	6	7	8	9	11	15	12	13	14

- **Description of a macroblock**

The macroblock is the basic unit in the MPEG stream, it is an area of 16x16 pixels and it is at this stage that the first compression takes place. Each pixel has a luminance (Y) component and two chrominance (Cb and Cr) components associated with it. The human eyes are much more sensitive to luminance than to chrominance. Therefore the luminance components must be encoded at full resolution while the chrominance components can be encoded at quarter resolution without any noticeable loss. This gives compression of 1:2 already. Figure 3.13 shows this compression.

A block is an 8 by 8 pixel area and is the smallest unit in the MPEG stream. It contains the DCT coefficients of the luminance and chrominance components (Rao and Hwang, 1996). Six blocks are needed to make up a macroblock (16 pixels by 16 pixels), four for the luminance components but only one for each of the two chrominance components due to their compression. Figure 3.14 shows the blocks of a macroblock and their numbering convention.

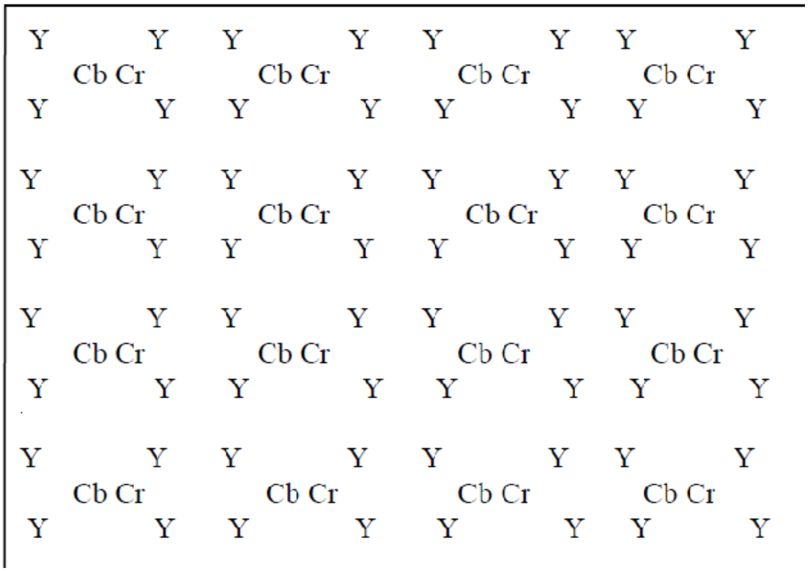

Figure 3.13 : Compression

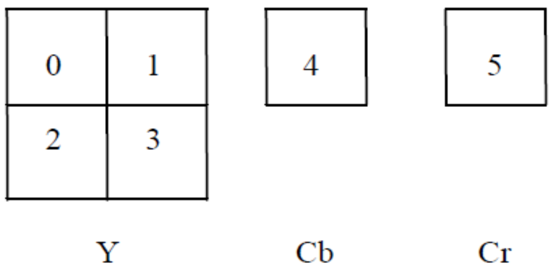

Figure 3.14 : Structure of a Macroblock and Block Numbering Convention

3.5.2. Types of Macroblock Present in a Frame

In a single frame there may be many different types of macroblock (MB). Tables 3.2, 3.3 and 3.4 show the different types of macroblock that can be present in I, P and B frames respectively.

TABLE 3.2

MACROBLOCK TYPES IN I FRAME

Type	VLC Code	MB Quantization
Intra - d	1	0
Intra - q	01	1

TABLE 3.3

MACROBLOCK TYPES IN P FRAME

Type	VLC	Intra	MF	Coded Pattern	Quant
pred-mc	1				
pred-c	01				
pred-m	001		1	1	
intra-d	0001 1		1	1	
pred-mcq	0001 0	1			1
pred-cq	0000 1		1	1	1
intra-q	0000 01	1		1	1
skipped					

In an I frame there are only two types of macroblock, Intra-d uses the default quantizer scale while Intra-q uses a scale defined by the buffer status. A P frame uses motion estimation and compensation to reduce the amount of information needed to play the video. There are eight different types of macroblock in a P frame, but for the purpose of this project they can be divided up into three categories.

1. Intra : There are no motion vectors present. These macroblocks don't use any reference frame, and are encoded using only information from itself.

2. Predicted : These macroblocks have motion vectors present.

3. Skipped : These macroblocks are the exact same as the macroblock in the previous frame.

TABLE 3.4

MACROBLOCK TYPES IN B FRAME

Type	VLC	Intra	MB	MF	Coded Pattern	Quant
pred-i	10		1	1		
pred-ic	11		1	1	1	
pred-b	010			1		
intra-bc	011			1	1	
pred-f	0010		1			
pred-fc	0011		1		1	
intra-d	0001 1	1				
pred-icq	0001 0		1	1	1	1
pred-fcq	0000 11		1		1	1
pred-bcq	0000 10			1	1	1
intra-q	0000 01	1				1
skipped						

In the above discussion, VLC is Variable Length Code, MF is Motion Forward, MB is Motion Backward, pred is predictive, m means motion compensated, c means at least one block in the macroblock is coded and transmitted, d means default quantizer is used, q means quantizer scale is changed, i means interpolated. This is a combination of forward prediction and backward prediction and b indicates backward prediction and f indicates forward prediction.

A B-frame uses two reference frames for prediction and can have twelve different types of macroblock. This leaves it the most complex, but it gives the highest compression rate. For the purpose of this research work, they can be categorized in five groups:

1. Forward predicted : Macroblock in encoded using only a past I or P frame.
2. Backward predicted : Macroblock in encoded using only a future I or P frame.
3. Forward and Backward predicted (Interpolated) : Macroblock is encoded using both a past and future frame as a reference. The two macroblocks are interpolated to form the predicted macroblock.
4. Intra : No reference frame is used. Macroblock is encoded using information from itself.
5. Skipped : Macroblock is the same as the one in the previous frame

3.5.3. Motion Estimation and Compensation

MPEG achieves high compression rate by the use of motion estimation and compensation. MPEG takes advantage of the fact that from frame to frame there is very little change in the picture (usually only small movements). For this reason macroblock size areas can be compared between frames, and instead of encoding the whole macroblock again the difference between the two macroblocks is encoded and transmitted.

Figure 3.15 demonstrates how forward motion compensation is achieved (backward compensation is done in the same way except a future frame in the display order is used as the reference frame).

Macroblock “x” is the macroblock we wish to encode, macroblock “y” is its counterpart in the reference frame. A search is done around “y” to find the best match for “x”. This search is limited to a finite area, and even if there is a perfectly matching macroblock outside the search area, it will not be used. The displacement between the two macroblocks gives the motion vector associated with “x”. There are many search algorithms to find the best matching macroblock. A full search gives the best match but is computationally expensive. Alternatives to this are the Logarithmic search, One-at-a-time search, Three-step search and the Hierarchical search. The choice of search is decided by the encoder, with the usual trade-off between time and accuracy.

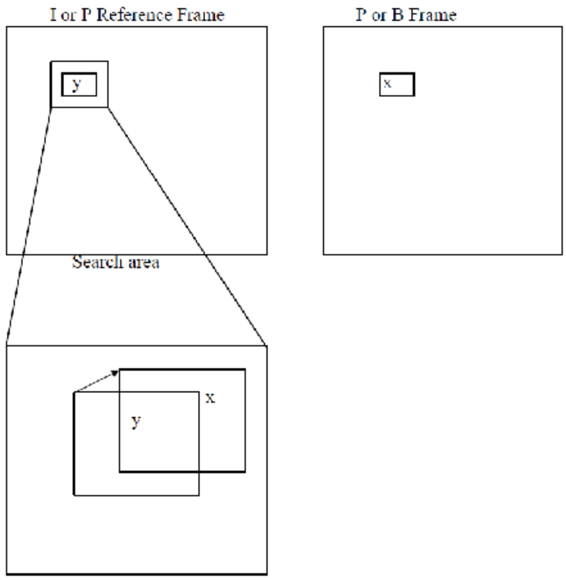

Figure 3.15 : A Forward Predicted Motion Vector

- **Encoding the Motion Vectors**

Once the motion vector is found it has to be encoded for transmission. The first step in the encoding process is to find the differential motion vectors (DMV). In a lot of situations, (e.g. a pan) all motion vectors will be nearly the same. Therefore subtracting the motion vector for a macroblock from the previous motion vector in the slice will reduce a lot of the vectors to zero. Note this differential vector is reset to zero if an I-type macroblock is encountered, and also at the end of a slice.

The second step is to make sure all differential vectors are within a permitted range. If the vectors are outside this range, a modulus is added/subtracted. Finally the differential vectors are variable length coded and transmitted.

3.6. CHAPTER SUMMARY

In this research work, watermarking techniques to protect video digital data are proposed. The study first identifies optimal frames and regions for embedding the watermark from uncompressed and compressed domains. Transformation based and feature based techniques are enhanced for this purpose. The following chapter (Chapter 4, Design of Frame and Regions Selection Algorithm) presents the steps involved during the search of optimal frames and regions for embedding copyright information in the video.