

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

185,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Compensation Methods for Video Coding

Ben-Shung Chow

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/53044>

1. Introduction

The standard dynamic image compression is usually composed of motion compensation and a DCT transformation for the error image from the motion compensation. The DCT coefficients are also classified by the cases to be coded by Huffman compression. However, the DCT coding will be inefficient for the binary images due to its broad dynamic range. It is noted that the binary images are usually described properly by their shapes. In this sense, a novel idea of shape compensation is proposed to replace the DCT processing. A schematic diagram to present this idea is illustrated in Fig. 1.

More clearly, the binary dynamic images are compressed by the shape compensation following the motion compensation. Our binary images are coded by the motion vectors and the kinds of shape transformations. For this transformation, a morphological filter is selected to modify the shape of the objects in image. Morphological processing [1] is a type of operation, by which the spatial form or structure of objects within an image are modified in a nonlinear way. The computation cost of nonlinear processing for conventional numerical image processing is very expensive. However, the morphology image processing treats the image components as sets and deals with the changes of shapes very efficiently. Thus, the morphology processing has recently been applied successfully to the industry auto-inspection and medical image processing [2-3]. The efficiency of morphological image processing makes the shape compensation in the decoding procedure very simple.

2. Conventional compensation method: Motion compensation

It is the speech transmission conventionally to be the major application in the real time communication. Thus, the basic rate in ISDN (Integrated Services Digital Network) was $2B+D =$

2*64+16 kbps in earlier days. Accordingly, there is no motive to propose video coding with fewer bit rates. As one example, video conference was developed under the ITU H.320 [4]. One exceptional research [5] works on the video coding to fit into the 20-40 kbps bandwidth range because this range can be stably provided in the 2.5 G wireless networks such as General Packet Radio Service (GPRS) and Code Division Multiple Access (CDMA) although the theoretical bandwidths of GPRS and CDMA 1X are 115 and 153.6 kbps respectively. In this mainstream, MPEG/H.26x [6] performs well in the bandwidth range about 64 kbps.

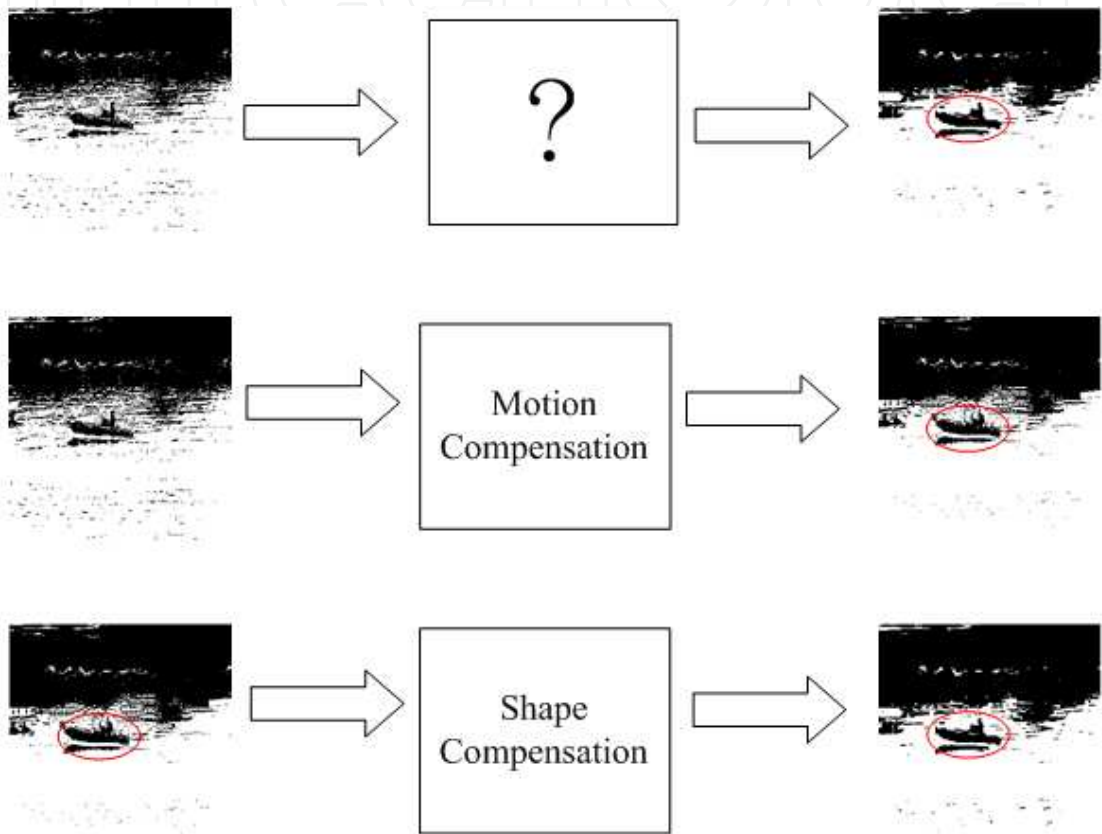


Figure 1. A schematic diagram to present the idea of shape compensation and comparison with motion compensation.

The MPEG digital video-coding techniques are statistical in nature. Video sequences usually contain statistical redundancies in both temporal and spatial directions. The basic statistical property upon which MPEG compression techniques rely is inter pixel correlation, including the assumption of simple, correlated motion between consecutive frames. Thus, it is assumed that the magnitude of a particular image pixel can be predicted from nearby pixels within the same frame (using intraframe coding techniques) or from pixels of a nearby frame (using interframe techniques). Intuitively, it is clear that in some circumstances, i.e., during scene changes of a video sequence, the temporal correlation between pixels in nearby frames is small or even vanishes. The video scene then assembles a collection of uncorrelated still images. In this case, intraframe coding techniques are appropriate to explore spatial correlation to achieve efficient data compression.

Motion in many video-conferencing scenes can be modeled to be primarily translational in nature. One popular procedure to estimate this type of motion is the block-matching technique [7]. Here the motion of a group of pixels (block) is represented by a single displacement vector, which is found by a matching technique. The so-called block motion-compensated coding schemes estimate the displacements of moving objects and only encode the block differences in moving areas between the current frame and the translated previous frame. Researchers have shown that this technique can increase coding efficiency significantly. A complete block motion-compensated coding system typically consists of three stages: (1) a motion detector which detects the moving blocks, (2) a displacement estimator which estimates the displacement vectors of moving blocks, and (3) a data compressor, which encodes the differences after motion compensation. Some terminology excerpted from GENERIC CODING OF MOVING PICTURES AND ASSOCIATED AUDIO, Recommendation H.262, ISO/IEC 13818-2 [8] are listed below for quick reference.

3. Terminology

Macroblock: The four 8-by-8 blocks of luminance data and the two (for 4:2:0 chrominance format), four (for 4:2:2 chrominance format) or eight (for 4:4:4 chrominance format) corresponding 8-by-8 blocks of chrominance data coming from a 16-by-16 section of the luminance component of the picture. Macroblock is sometimes used to refer to the sample data and sometimes to the coded representation of the sample values and other data elements defined in the macroblock header of the syntax defined in this part of this specification. The usage is clear from the context.

Motion compensation: The use of motion vectors to improve the efficiency of the prediction of sample values. The prediction uses motion vectors to provide offsets into the past and/or future reference frames or reference fields containing previously decoded sample values that are used to form the prediction error signal.

Motion estimation: The process of estimating motion vectors during the encoding process.

Motion vector: A two-dimensional vector used for motion compensation that provides an offset from the coordinate position in the current picture or field to the coordinates in a reference frame or reference field.

Non-intra coding: Coding of a macroblock or picture that uses information both from itself and from macroblocks and pictures occurring at other times.

- **P-field picture:** A field structure P-Picture.
- **P-frame picture:** A frame structure P-Picture.
- **P-picture; predictive-coded picture :** A picture that is coded using motion compensation.

With the above terminology, some illustrating examples and schematic diagrams are presented below to explain the details of motion compensation in MPEG.

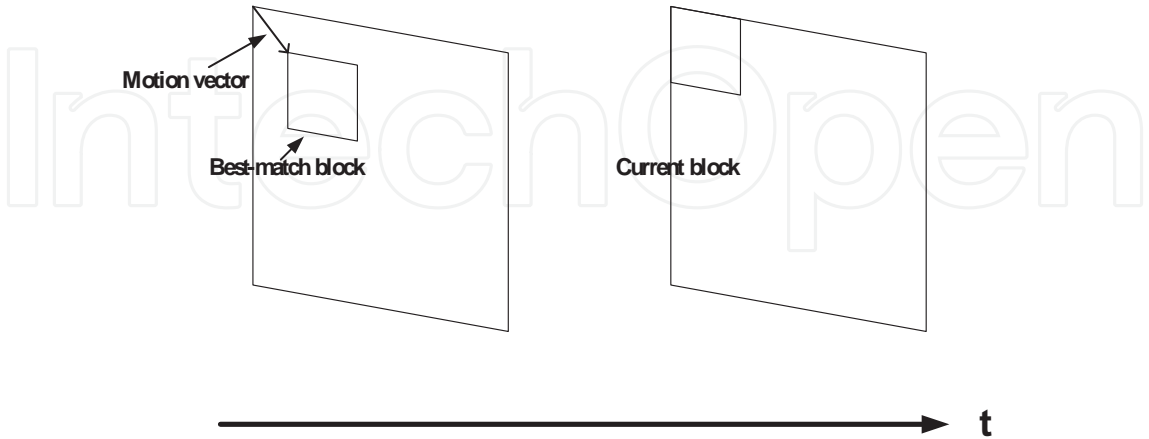


Figure 2. Illustration example for forward direction motion vector.

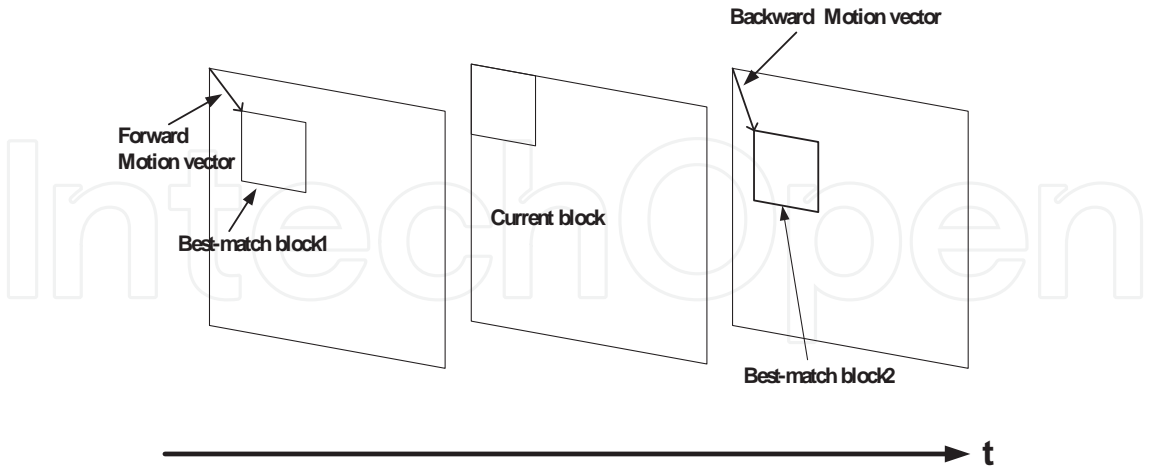


Figure 3. Illustration example for bi-direction motion vector.

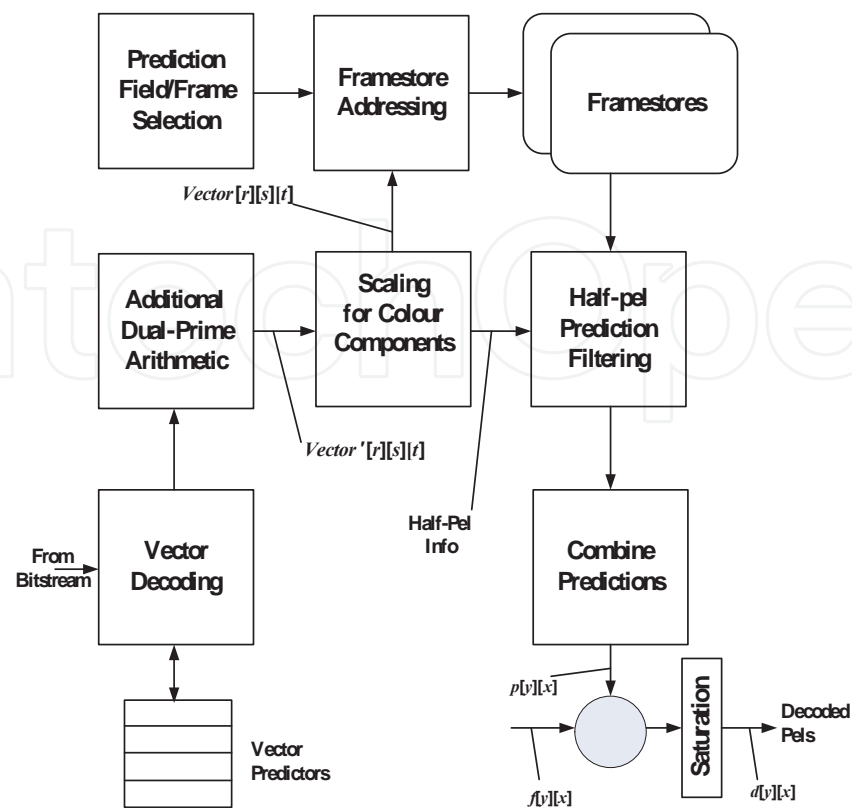


Figure 4. Flow chart for MPEG coding procedures.

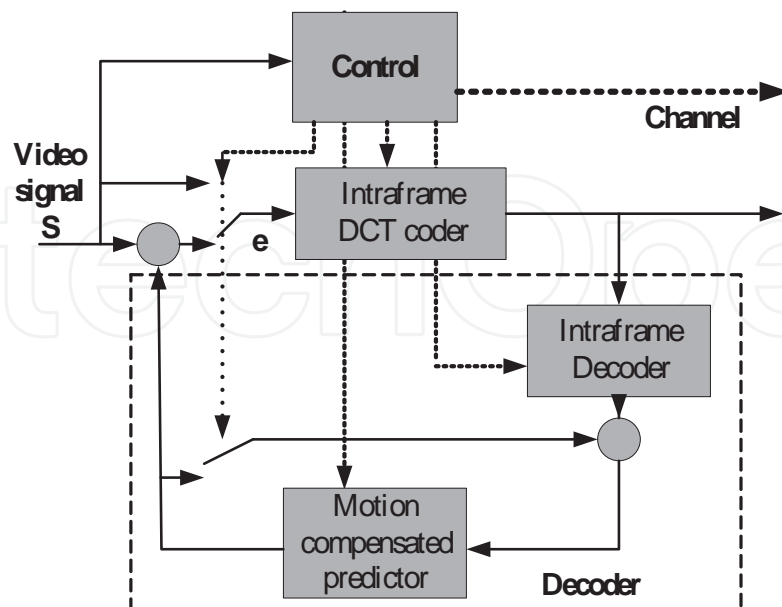


Figure 5. Schematic diagrams for MPEG coding procedures.

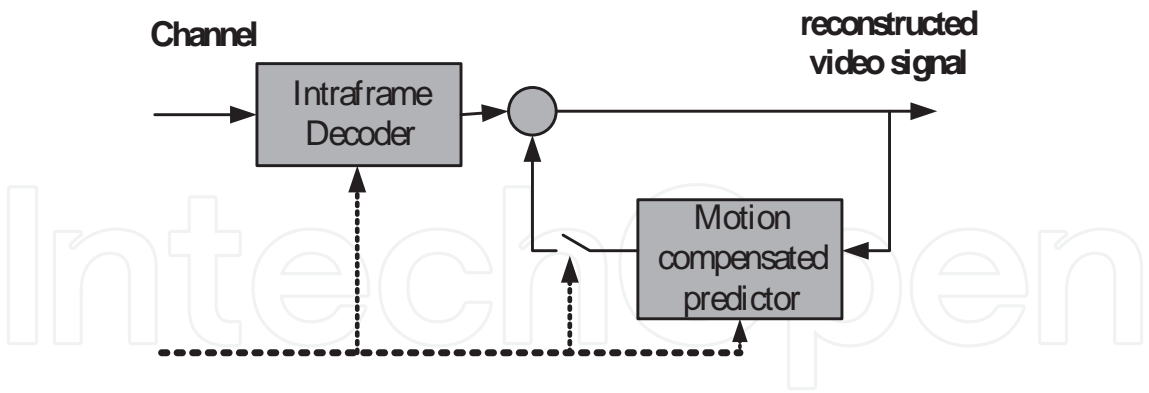


Figure 6. Schematic diagrams for MPEG decoding procedures.

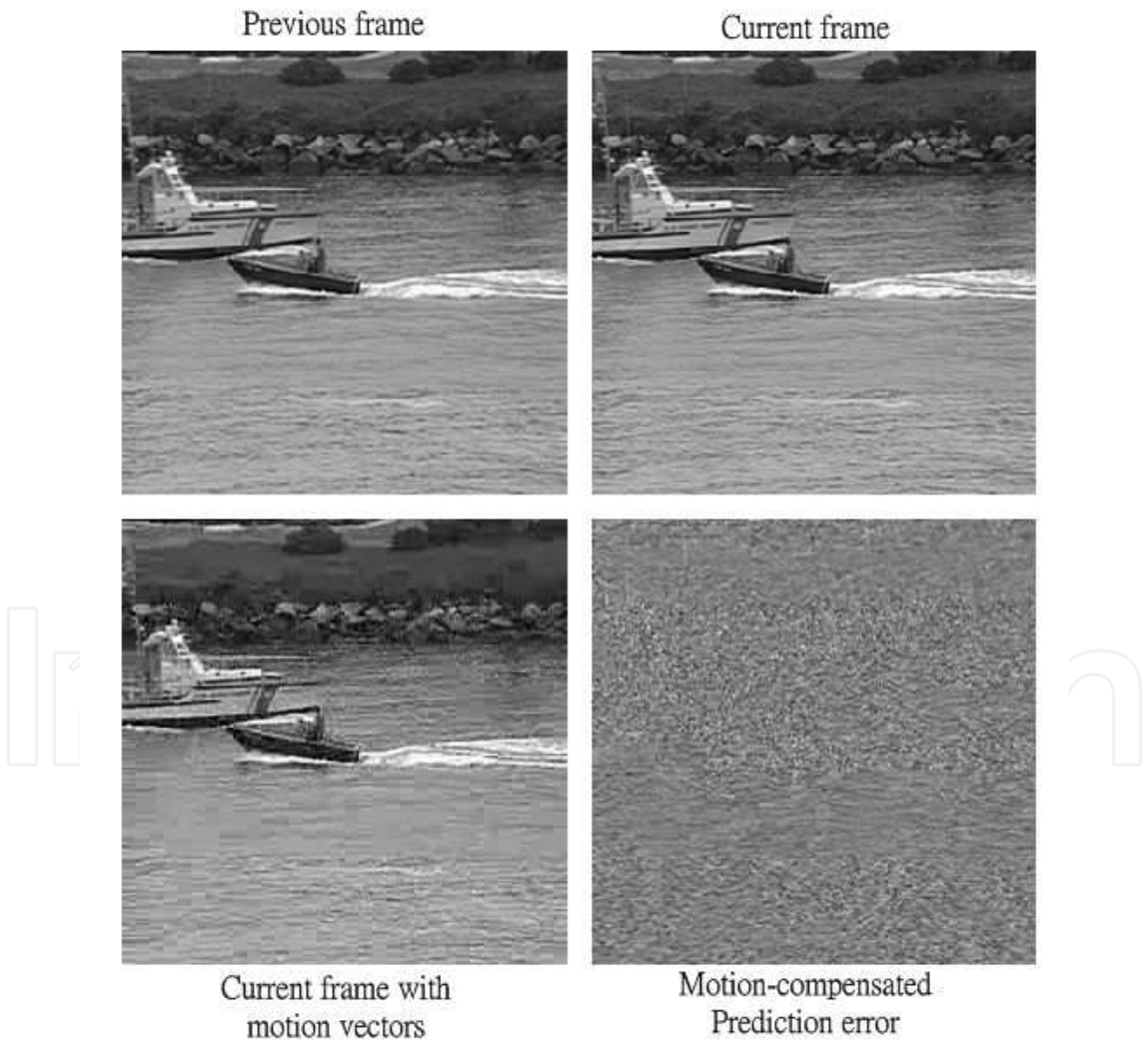


Figure 7. Image examples to illustrate motion compensation (a) the reference image (b)the coding image (c) the motion compensated image (d) the error image between b and c.

4. Shape transform by mathematical morphology

Mathematical morphology refers to a branch of nonlinear image processing and analysis developed initially by Serra [1] that concentrates on the geometric structure within an image. The morphological approach is generally based upon the analysis of a two value image with some predetermined geometric shape known as a structuring element. The basic idea, arising out of stereology, is to probe an image with a structuring element and to quantify the manner in which the structuring element fits or does not fit within the image. The analysis of the geometric objects must be quantitative, because only such an analysis and description of geometric objects can provide a coherent mathematical framework for describing the spatial organization. The quantitative description of geometrical structures is the purpose of mathematical morphology.

The scope of morphological method is as wide as image processing itself. These include enhancement, segmentation, restoration, edge detection, texture analysis, particle analysis, feature generation, skeletonization, shape analysis, compression, component analysis, curve filling, and general thinning. There are many areas where morphological methods have been successfully applied, including robot vision, inspection, microscopy, medical imaging, remote sensing, biology, metallurgy, and automatic character reading.

The language of mathematical morphology is set theory. As such, morphologies offer a unified and powerful approach to numerous image processing problem. Sets in mathematical morphology represent the shapes of objects in an image. In binary images, the sets in question are members of 2-D integer space Z^2 , where each element of a set is a tuple(2-D vector) whose coordinates are the (x,y) coordinates of image object pixel(foreground by convention) in the image.

In the following sections we illustrate several important concepts in mathematical morphology, and a detailed discussion can be found in [1].

4.1. Dilation and erosion

The simplest morphological transformation is dilation and erosion. In loose terms, these operations cause the swelling or the shrinking of areas when structuring element has a disklike shape. They are the base for most of the morphological operations discussed later.

Some basic definitions

Let A and B be sets in Z^2 , with components $a=(a_1, a_2)$ and $b=(b_1, b_2)$, respectively. The translation of A by $x=(x_1, x_2)$, denoted, is defined as

$$(A)_x = \{c \mid c = a + x, \text{ for } a \in A\} \quad (1)$$

The reflection of B, denoted , is defined as

$$\overset{\vee}{B} = \{x \mid x = -b, \text{ for } a \in B\} \quad (2)$$

The complement of set A is

$$A^c = \{x \mid x \notin A\} \quad (3)$$

4.1.1. Dilation

With image A and a structuring element B as sets in Z^2 , the dilation of A by B, denoted $A \oplus B$, is defined as

$$A \oplus B = \{y \mid y = a + b, a \in A, b \in B\} \quad (4)$$

There are two equivalent formulations of dilation that deserve. First,

$$A \oplus B = \bigcup_{b \in B} (A)_b \quad (5)$$

So that the dilation can be found by translating the input image by all points in the structuring element and then taking the union. Fig 2.1 illustrates the concept.

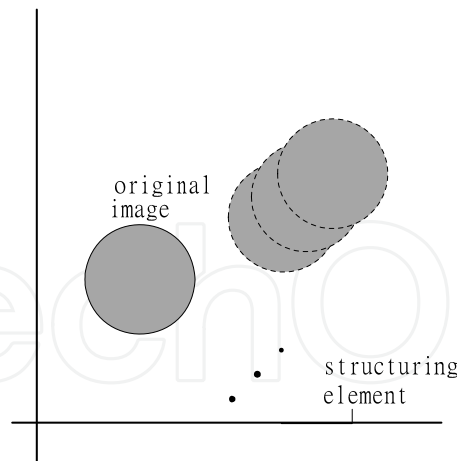


Figure 8. Dilation as a union of translates

Another formulation of dilation involves translates of the rotated structuring element that "hit" (intersect) the input image.

$$A \oplus B = \{x \mid (\overset{\vee}{B})_x \cap A \neq \emptyset\} \quad (6)$$

Thus the dilation process consists of obtaining the reflection of B about its origin and then shifting this reflection by x . The dilation of A by B then is the set of all x displacements such that B and A overlap by at least one nonzero element.

4.1.2. Erosion

For sets A and B in Z^2 , the erosion of A by B , denoted $A \ominus B$, is defined as

$$A \ominus B = \{x \mid x + b \in A \text{ for every } b \in B\} \quad (7)$$

Equivalently, we may write

$$A \ominus B = \bigcap_{b \in B} (A)_{-b} \quad (8)$$

Here, the erosion is found by intersecting all translates of the input image by negatives of points in the structuring element. The method is illustrated in Fig 2.2.

Another definition of erosion is the following:

$$A \ominus B = \{x \mid (B)_x \subseteq A\} \quad (9)$$

Which, in words, says that the erosion of A by B is the set of all point x such that B , translate by x , is contained in A . If the origin lies inside the structuring element, then erosion has the effect of shrinking the input image. Geometrically, the disk B has been moved around inside of A and the positions of the origin has been marked to produce the eroded image. Formally, we can state the following property: if the origin is contained within the structuring element, then the eroded image is a subset of the input image.

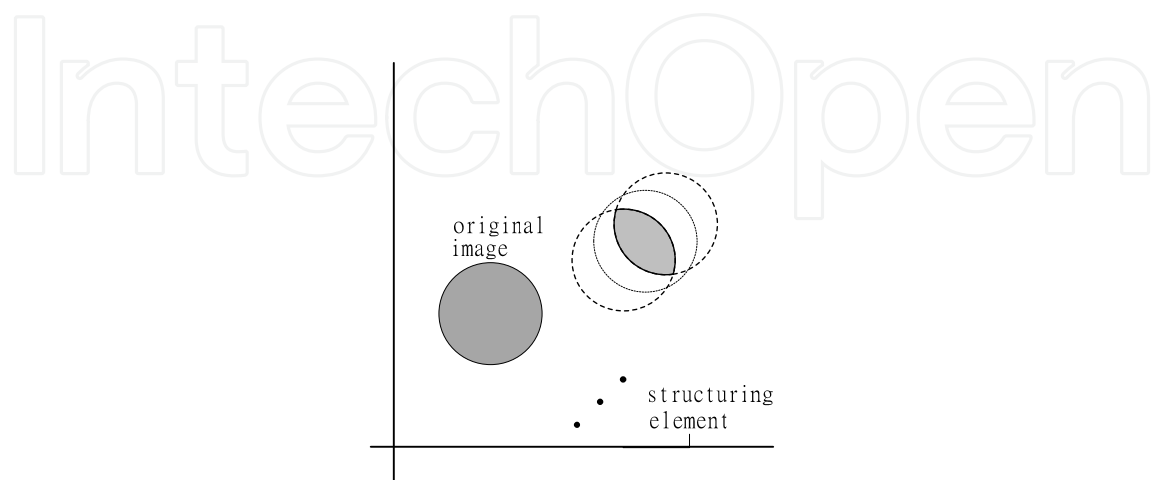


Figure 9. Erosion as an intersection of translates

Dilation and erosion are duals of each other with respect to set complementation and reflection. That is,

$$A \ominus B = (A^c \oplus B)^c \quad (10)$$

$$A \oplus B = (A^c \ominus B)^c \quad (11)$$

proof: Let $z \in (A^c \ominus B)^c$. Then $z \notin A^c \ominus B$. This happens if and only if there exists a $b \in B$ such that $z \notin A_b^c$. This also happens if and only if there exists a $b \in B$ such that $z \in A_b$, i.e. $z \in A \oplus B$.

If we think of a disk structuring element, then dilation fill in small (relative to the disk) holes and protrusions into image, whereas erosion eliminates small components and extrusions of the image into its complement. Here, we only discuss the implementation of dilation instead of erosion. Applying the duality property of dilation and erosion, we can easily employ both logical and geometric negation to implement erosion because of the different roles of the image and the structuring element in an expression employing these morphological operations.

5. Application of Matheron representation theorem

Matheron representation theorem with the notion of kernel and basis are the fundamental theory for morphological filter applied to video coding. However, the theory of representation of morphological filter is usually lengthy in textbooks [1]. We therefore have a brief introduction here.

The basic image operations for morphology processing are dilation and erosion. Erosion is the dual of dilation. They are totally built upon the Boolean operations: set union, intersection and complementation. The result of the dilation is the union of all translates of the original image. Every translation distance is determined by the coordinate of the corresponding black pixel in another image called the structuring element.

Definition 1: increasing

A mapping $\Psi : 2^{R^*R} \rightarrow 2^{R^*R}$ is increasing if whenever $A \subset B$, then $\Psi(A) \subset \Psi(B)$.

The definition of decreasing can be similarly defined.

Definition 2: translation invariant

A mapping Ψ is translation invariant if $\Psi(T(A)) = T(\Psi(A))$ for all $A \in 2^{R^*R}$, where T is the translation mapping.

Definition 3: kernel

The kernel of an increasing translation invariant mapping Ψ , denoted by $\text{Ker}[\Psi]$, is important to the theory of morphological filters and is defined by the collection of all the images that will contain the origin pixel (0, 0) mapped by that translation invariant mapping Ψ .

The set of kernel can be further reduced to the basis defined below.

Definition 4: basis

A basis for the Kernel of an increasing translation invariant mapping is defined by satisfying the following two conditions:

1. No element of this basis is a proper subimage of any other element of this basis
2. For any element T belonging to the corresponding kernel, there exists an element T' of this basis such that T' is a subset T .

Theorem 1: Matheron Representation Theorem 1

Any increasing (decreasing) translation invariant morphological filter can be expressed as a union (intersection) of erosions (dilation) over sets of its kernel as follows

$$\Psi(A) = \bigcup_{B \in \text{Ker}[\Psi]} \text{Erosion}(A, B).$$

Theorem 2: Matheron Representation Theorem 2

$$\Psi(A) = \bigcup_{B \in \text{Bas}[\Psi]} \text{Erosion}(A, B), \text{ where } \text{Bas}[\Psi] \text{ is the basis of } \text{Ker}[\Psi].$$

From Matheron representation theorem, every decreasing (increasing) translation invariant image mapping can be composed of a union (join) of erosion (dilation) operations. According to this theorem, the idea of kernel for the above mentioned increasing mapping is introduced by referring the set of the structuring elements of those erosions or dilations to characterize that mapping. Furthermore, to simplify the computation of the above mapping, the concept of the basis for kernel is further introduced, borrowing the idea of the basis from linear space, to reduce the number of the involved structuring elements.

As a summary, the morphological mapping is characterized by its kernel, and the kernel is characterized by its basis. Consequently, the proposed morphological transform (shape compensation) is coded by the kinds of basis or morphological filter.

6. Shape compensation: Selection of morphological filters

The concept of shape compensation is implemented on two image blocks: source block and target block. The source block is shape compensated (filtered) to look like the target block. The frames are divided into blocks of the size of 16*16. We need to select a filter for each source block. We first define the pattern of a cross form on the source block. In the proposed coding method, the source block is the motion compensated previous coded block and the target block is the current coding block.

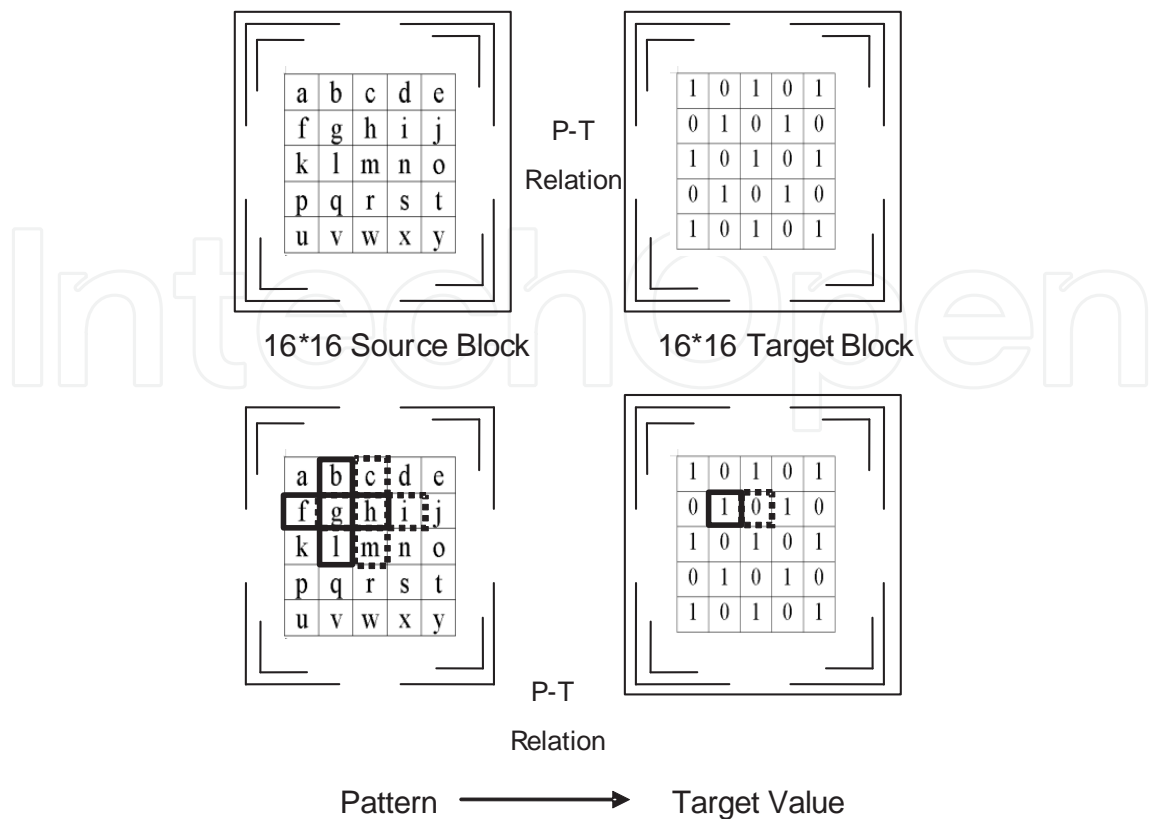


Figure 10. Pattern-Target relation illustration: pattern block and target values are scanned through the source block and the target block respectively; 3*1 mask is used in this illustrative example.

We then define the target pixel on the target block by the corresponding pixel with a same center location for the pattern on the source block. For each pattern, the filtered result is either consistent with the target pixel or not. The optimal filter is the filter causing the least inconsistent results. The optimal filter can be selected pattern by pattern. To reduce computation, the filter is selected group by group. The group is the pattern group classified by its kind of pattern. There are only 32 groups in the 14*14 patterns from the source block. The pattern group associated with the target value is called pattern-target relation or pattern-target occurrence table. One realization of this relation is shown in Table 1.

In practice, we first scan the source and target blocks to have this occurrence table. We next filter every group to have the pattern-filter relation table as in Table 2. It is noted that this table can be prepared in advance. One typical table composed of erosion filters is shown in Table 2. Finally, we build a pattern-filter conflict table from the pattern-target relation (pattern-target occurrence table) and the pattern-filter relation table. Using this table, the least inconsistent filter is obtained. One example of the pattern-filter conflict table deduced from Table 1 and Table 2 is shown in Table 3. A schematics diagram to introduce the relation between the pattern and target pixel is illustrated in Fig. 2. It is noted that the conflict count is produced in a single scan of the block pairs in the source and target image frames to save time.

Pattern-Target Relation					
Pattern Type	Target Value	Occurrence#	Pattern Type	Target Value	Occurrence#
00000	0	65	10000	0	0
	1	1		1	0
00001	0	1	10001	0	0
	1	0		1	0
00010	0	1	10010	0	0
	1	1		1	0
00011	0	2	10011	0	0
	1	0		1	1
00100	0	0	10100	0	0
	1	0		1	0
00101	0	0	10101	0	0
	1	0		1	0
00110	0	1	10110	0	0
	1	0		1	0
00111	0	0	10111	0	0
	1	1		1	1
01000	0	10	11000	0	0
	1	0		1	0
01001	0	2	11001	0	1
	1	1		1	0
01010	0	0	11010	0	0
	1	0		1	0
01011	0	0	11011	0	0
	1	0		1	0
01100	0	0	11100	0	1
	1	0		1	0
01101	0	3	11101	0	3
	1	1		1	6
01110	0	0	11110	0	0
	1	0		1	2
01111	0	0	11111	0	28
	1	1		1	62

Table 1. One example of pattern-target relation

Filter	Filter 00000	Filter 00001	Filter 00010	Filter 11111
Pattern				
Pattern Type	Filtered Value	Filtered Value	Filtered Value	Filtered Value
00000	0	0	0	0
00001	0	0	0	1
00010	0	0	0	1
00011	0	0	0	1
00100	0	0	0	0
00101	0	0	0	1
00110	0	0	0	1
00111	0	0	0	1
01000	0	0	0	0
01001	0	0	0	1
01010	0	0	0	1
01011	0	0	0	1
01100	0	0	0	0
01101	0	0	0	1
01110	0	0	0	1
01111	0	0	0	1
10000	0	0	0	0
10001	0	0	0	1
10010	0	0	0	1
10011	0	0	0	1
10100	0	0	0	0
10101	0	0	0	1
10110	0	0	0	1
10111	0	0	0	1
11000	0	0	0	0
11001	0	0	1	1
11010	0	0	0	1
11011	0	1	1	1
11100	0	0	0	0
11101	1	0	1	1
11110	1	1	1	1
11111	1	1	1	1

Table 2. One example of pattern-filter relation

Filter Pattern	Filter 00000	Filter 00001	Filter 00010	Filter 11111
Pattern Type	Occurrence #	Occurrence #	Occurrence #	Occurrence #
00000	1	1	1	1
00001	0	0	0	1
00010	1	1	1	1
00011	0	0	0	2
00100	0	0	0	0
00101	0	0	0	0
00110	0	0	0	1
00111	1	1	1	0
01000	0	0	0	0
01001	1	1	1	2
01010	0	0	0	0
01011	0	0	0	0
01100	0	0	0	0
01101	1	1	1	3
01110	0	0	0	0
01111	1	1	1	0
10000	0	0	0	0
10001	0	0	0	0
10010	0	0	0	0
10011	1	1	1	0
10100	0	0	0	0
10101	0	0	0	0
10110	0	0	0	0
10111	1	1	1	0
11000	0	0	0	0
11001	0	0	1	1
11010	0	0	0	0
11011	0	0	0	0
11100	0	0	0	0
11101	3	6	3	3
11110	0	0	0	0
11111	28	28	28	28
Total Conflict	39	42	40	43

Table 3. One example of pattern-filter conflict

As an illustrating example, two 16×16 block obtained from Fig. 14(a) and Fig. 13(b) in the same location are used for the data in Table 1, 2, and 3. In the source block from Fig. 14(a), there are 66 occurrences of pattern 00000, while 65 of which are aimed to be 0 and 1 is aimed to be 1 according to the target block in Fig. 13(b) as shown in the first square in Table 1. Next, there are 32 selected filaters, each of which are composed of two erosion masks. The first filter 00000, composed of the masks 11110 and 11101, will filter the pattern 00000 to be the value 0 as shown in the first square in Table 2. Thus, there is 1 occurrence of pattern 00000 that will not favor the action of filter 00000 since this filter will change it to the value 0 (as in Table 2) but 1 of this pattern is targeted to 1 (as in Table 1). Consequently, the occurrence of conflict between the pattern 00000 and the filter 00000 is 1 as shown in the first square in Table 3.

7. Experimental results

The video streams of the “boat” are tested in our experiment. The original streams are of the CIF format with the gray levels of 256. They are first to be reduced to the size of 256×256 , then threshold to be bi-level 256×256 sequences. Binary image processing is thus applied to the bi-level sequences in the encoding stage, which includes the motion vectors search and the optimal morphological filters determination. The block would be refreshed if the motion compensated error is above the threshold. Correspondingly, there are two possibilities in the decoding process: block refreshing or shape compensation after motion compensation.

For the purpose of comparison, two coding schemes are arranged. The first one is the conventional coding by motion compensation only. The second is coding with motion compensation followed by shape compensation. The frame errors of the three coding methods for the entire boat streams are plotted in Fig. 11. For visual quality demonstration, the first frame of the streams is shown with the forms of grey-leveled 256×256 in Fig. 12(a) and the form of bi-leveled are shown in the Fig. 12(b). Finally, the image samples of the decoded streams from coding one and two are illustrated in the Fig. 13 and Fig. 14 with being magnified to see the detail in Fig. 15.

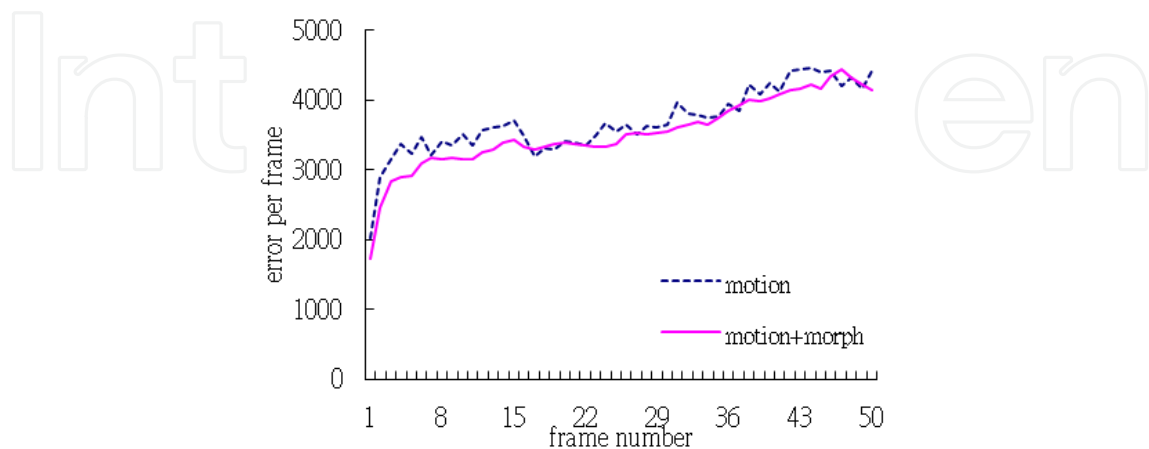


Figure 11. Error comparison to illustrate the error improvements by using the proposed shape compensation after motion compensation.



Figure 12. Frame 1 in the form of 256*256 (a) gray levels (b) bi-level for visual quality demonstration.



Figure 13. Frame 13 in the form of 256*256 (a) gray levels (b) bi-level.



Figure 14. a) The conventional decoded (b) the decoded by motion and shape compensation.



Figure 15. a) The conventional decoded (b) the decoded by motion and shape compensation, magnified to see the detail.

8. Conclusions

Our coding procedures are simple: binary coding by motion compensation and shape compensation without using DCT coding. In the encoding stage, every motion compensated block has a shape compensation by a suitable morphological filter. This filter is selected on-line from a set of filters, which is selected off-line based on known statistics and experiences. The selection is by voting strategy. Decoding is simply a motion compensation followed by an efficient morphological filtering.

Visual quality for human is not concerned in this method as in the cases of telephony or conferences. In some situations such as the wild filed monitoring system, the decision function of security systems can be shifted to the concerned human if the video information can be provided inexpensively. For instance, the existence or the location of the intruder, which is the key information should be provided to the monitoring system. The low visual quality is tolerated with the convenient help of human decision-making as presented in this chapter.

Acknowledgements

This research is partially supported by the National Science Council of the Republic of China under the contract NSC 99-2625-M-009-004-MY3

Author details

Ben-Shung Chow

Address all correspondence to: bschow@ee.nsysu.edu.tw

Department of Electrical Engineering, National Sun Yat-Sen University, Lienhai Rd., Kaohsiung, Taiwan, R.O.C.

References

- [1] J. Serra . Image Analysis and Mathematical Morphology. New York: Academic Press, 1983.
- [2] J. W. Klingler, C. L. Vaughan, and L. T. Andrews, "Segmentation of echoboarddiographic images using mathematical morphology, " IEEE Trans. on Biomedical Engineering, vol.35, No.11, pp. 925-933, 1988.
- [3] J. M. Higgins, D. T. Eddington, S. N. Bhatia, and L. Mahadevan1, "Statistical dynamics of flowing red blood cells by morphological image processing," PLoS Comput Biol., 5(2): e1000288. 2009.
- [4] G. J. Sullivan, and T. Wieland, "Video Compression- From concept to the H.264/AVC standard," Proc IEEE, vol. 93, pp. 18-35, Jan. 2005.
- [5] J. Li, K. Yu, T. He, Y. Lin, S. Li, and Q. Zhang, "Scalable portrait video for mobile video communication," IEEE Trans. Circuits Syst. Video Technol., vol. 13, no. 5, pp. 376–384, May 2003.
- [6] B. G. Haskell, P. G. Howard, Y. A. LeCun, A. Pury, J. Ostermann, M. R. Civanlar, L. Rabiner, L. Bottu, and P. Haffner. "Image and video coding - emerging standards and beyond, " IEEE Trans. Circuits. Syst. Video Technol., vol. 8, no. 7, pp. 814-837, Nov. 1998.
- [7] Jian Feng, Kwok-Tung Lo, Mehrpour, H., Karbowiak, A.E . Adaptive block matching motion estimation algorithm using bit-plane matching. Proceedings., 1995 International Conference on Image Processing, pp. 496 – 499.
- [8] MPEG-2 Standard. ISO/IEC Document 13818-2. Generic Coding of Moving Pictures and Associated Audio Information, 1994.

IntechOpen

