THREE DIMENSIONAL VIDEO SIGNAL PROCESSING AND ITS APPLICATION

Jing Fang
M.E., Tsinghua University, Beijing, China, 1987

A thesis submitted to the faculty
of the Oregon Graduate Center
in partial fulfillment of the
requirements for the degree
Master of Science
in
Electrical Engineering

December, 1989

The thesis "Three dimensional video signal processing and its application" by Jing Fang has been examined and approved by the following Examination Committee:

Ajay Luthra
Adjunct Assistant Professor
Thesis Research Advisor

J. Fred Holmes
Professor
Examination Committee Chairman

V. Rao Gudimetla
Assistant Professor

ACKNOWLEDGEMENT

I would like to express my deepest gratitude to my advisor, Dr. Ajay Luthra for his attention, his guidance and his encouragement throughout this work. I am also grateful to Dr. Stephen Kahne for his support and encouragement. Thanks are due to the member of the thesis committee Dr. J. Fred Holmes and Dr. V. Rao Gudimetla for their valuable suggestions and comments.

Last but not least, I wish to thank Beverly Kyler, the library staff and my friends for all their help.

To Nanrong Xu and Richard A. Elliott for their encouragement

Table of Contents

Table of Figures

ABSTRACT

Three Dimensional Digital Video Signal Processing
and Its Application

Jing Fang, M.S.
Oregon Graduate Center, 1989

Supervising Professor: Ajay Luthra

High definition television (HDTV) is expected to be the next generation video medium because it can provide a high quality picture. One of the key technologies for HDTV is motion estimation and motion compensation. In this thesis, 3-D spectra and 3-D filters in TV are addressed; current advanced television (ATV) and HDTV proposals are introduced; and a new motion estimation system suitable for HDTV is proposed and described, in which simple detection, block processing, optic flow estimation and correlation search make the estimation more accurate with less computational load.

INTRODUCTION

Black and white television was a practical public service nearly a half of a century ago. Although it has had a surprisingly long popularity considering its stunted representation of reality, it was not satisfactory as it was bereft of color. In the early 1950's, color television was invented. It made people comfortable since it resembled our real world. Now, a demand for getting a more realistic viewing experience than is offered by today's color television system breaks the era of High Definition TeleVision (HDTV).

Three big competitors in promoting the development of HDTV are Japan, Europe and the United States. In June 1989, Japan started daily hourlong experimental HDTV satellite broadcasting of its MUSE (MUltiple Sub-Nyquist Encoding) system to prototype HDTV sets scattered throughout the country and plans to increase it to eight hours per day when a new satellite is launched in 1990. In Europe, the Eureka 95 project --- HD-MAC (High Definition Multiplexed Analog-Component) is scheduled to be on satellite broadcasts experiment in 1991. The United States meanwhile is bogged down in disputes over a terrestrial transmission standard to be receivable by current NTSC (National Television System Committee) receiver albeit with less than HDTV quality. It is hoped that the standard will be selected in 1992 after testing all candidate systems and the

HDTV broadcasting will be experimented in 1995. So the United States has a lot of catching up to do.

When Japan launched the HDTV era in June, the HDTV came into our life. A so far uniquely Japanese application of HDTV is the exhibition of still picture in an art gallery. It is predicted that HDTV will be used not only for broadcasts but also for making movies, printing, publishing, medical imaging, merchandising and education.

HDTV is to make a good use of human visual and aural system to describe the real world in facsimile. The aim of HDTV includes enlarging the aspect ratio or getting widescreen presentation on a display, increasing the resolution both in horizontal direction and vertical direction, improving the picture quality such as suppression of flickering, and enhancing audio quality. Compatibility[1] with today's systems, marketability[2,82] for attracting consumers and manufacturers, a single international standard etc. have become hot topics recently. It is envisioned that the future environment for HDTV delivery to consumers will have to include all of the elements shown in Fig.1[1] to accommodate various national differences in HDTV standards.

With the advent of HDTV, more video signal processing is required than even the most advanced system available today. The HDTV signal requires a bandwidth of 30 MHz. In comparison, today's NTSC system requires only a 6 MHz bandwidth. It will not be possible to transmit a full 30 MHz bandwidth over the air. Therefore, a lot of attention is paid to data compression and data

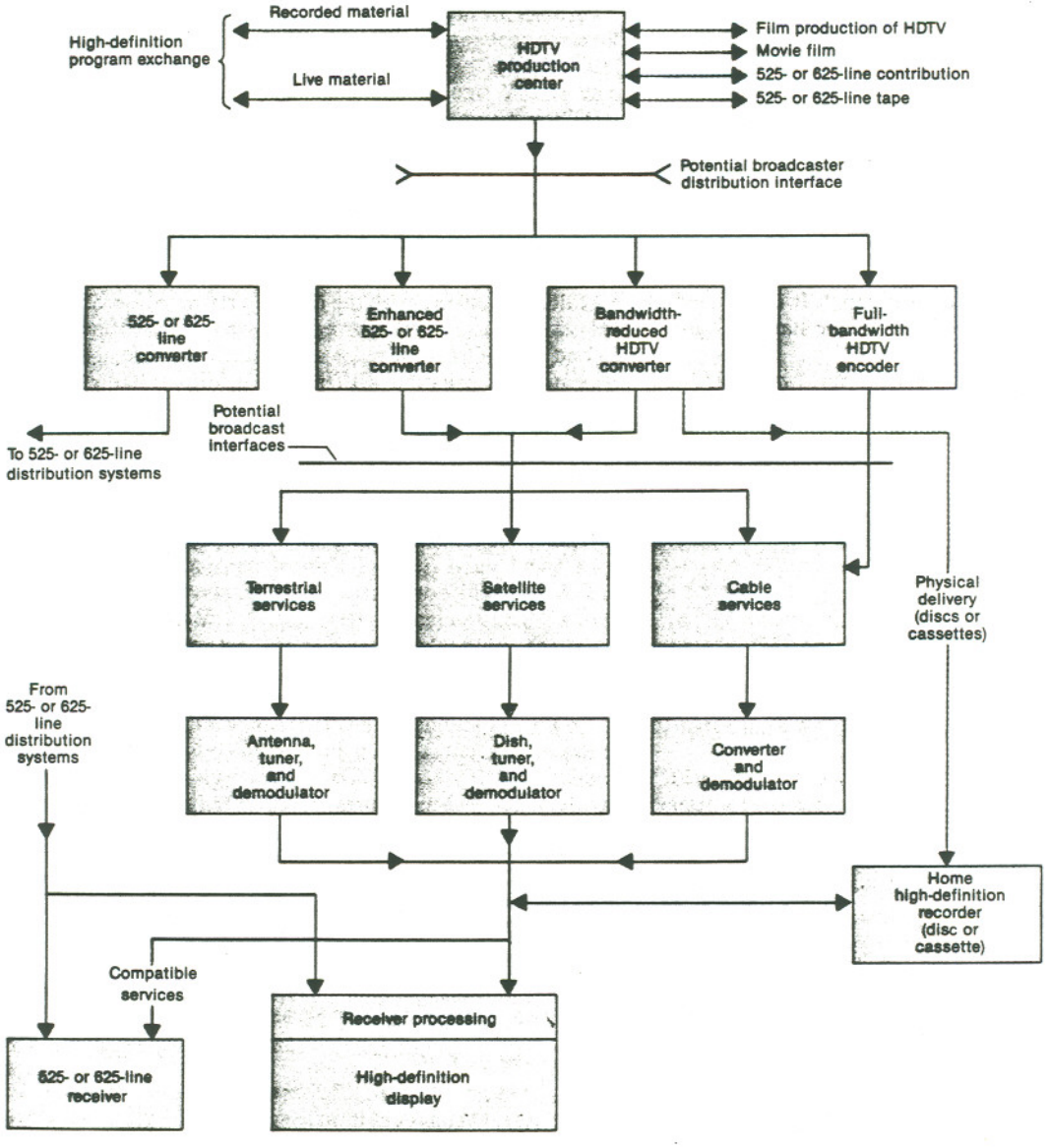


Fig.1 Block diagram of HDTV system in the future

recovery. Basically, the motion picture has more troubles than the stationary picture for signal processing. That is why the motion compensation is required in image compression and regeneration. The key point of effective motion compensation is motion detection and motion estimation. To improve the method of motion estimation and to make it as a practical one is what we want to do in this project.

Although a video signal can be described in one dimension, to analyze it only in one dimension is not sufficient and efficient even if in today's television systems. However, in the literature, there is no systematic analysis of the video signals in multi-dimensions. Therefore, another project in this thesis is to completely analyze the video signal in two and three dimensions.

In this thesis, we start from basic concepts and finally reach an improved motion estimation method. Chapter 1 gives an introduction of HDTV and this project. Chapter 2 covers an outline of concepts of three dimensional frequencies and three dimensional spectra which play a pivotal role in image analysis and image processing. In the beginning of Chapter 3, some common defaults are introduced and the method for avoiding these defaults are described. Then several HDTV proposals are introduced in brief. In Chapter 4, prominent three dimensional filters used in HDTV are studied. In Chapter 5, motion estimation methods are discussed and an improved estimation method is proposed.

THREE DIMENSIONAL SPECTRUM USED IN IMAGE PROCESSING

In multidimensional system (M-D and M≥2), the signal can be modeled as a function of M independent variables. These signals can be classified as continuous, discrete or mixed (discrete-continuous) signal. Usually, a scene is a three dimensional (3-D) signal since its gray value is defined on a set of points at certain time. Spectral analysis is the most powerful mathematical tool used for their analysis. In this chapter, we start by explaining the concept of frequency in image processing. Then, we will describe some special sequences and their spectra. Finally, we discuss NTSC (National Television System Committee) and PAL (Phase Alternation Line) TV (TeleVision) spectrum.

2.1. Concept of Frequency

A picture is described by the light intensity. The intensity is usually considered as a continuous function i of x, y and t, where x is an independent variable in the horizontal direction, y is an independent variable in the vertical direction and t is time. The spectrum of such a function can be represented by a three-dimensional (3-D) Fourier transform as described in Eq.(2.1).

Fourier Transform:

$$I(f_x,f_y,f_t) = \mathcal{F}\left[i(x,y,t)\right]$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} i(x,y,t)\, e^{-j2\pi(f_x x + f_y y + f_t t)} dx dy dt \qquad (2.1a)$$

where f_x, f_y, and f_t are the frequencies along the horizontal, the vertical and the temporal direction and are explained later.

Inverse Fourier Transform:

$$i(x,y,t) = \mathcal{F}^{-1}\left[I(f_x,f_y,f_t)\right]$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I(f_x,f_y,f_t)\, e^{j2\pi(f_x x + f_y y + f_t t)} df_x df_y df_t \qquad (2.1b)$$

The components of the 3-D spectrum are spoken of as "spatio-temporal frequencies". There are three kinds of frequences. The vertical frequency f_y, representing the change in intensity along y direction, is expressed in cycles per picture height (c/ph). The horizontal frequency f_x, representing the change in intensity along x direction, is represented by cycles per picture width (c/pw). In addition to these spatial frequencies, there is the temporal frequency f_t, which represents the change in the scene with time and is described in Hz. For simplicity, consider the black and white stationary images. In Fig.2.1(a), picture consists of a constant intensity. There are no changes in x or y direction with f_x=0 and f_y=0. If there are changes only in horizontal direction as Fig.2.1(b) with f_x=1 and f_y=0 and Fig.2.1(c) with f_x=2 and f_y=0, strips are perpendicular to the horizontal direction. If f_x and f_y both are non zeros, such as Fig.2.1(e), the strips will incline. Also, from Fig.2.1, we can find that the diagonal frequency in one direction is the summation of vertical and horizontal frequencies

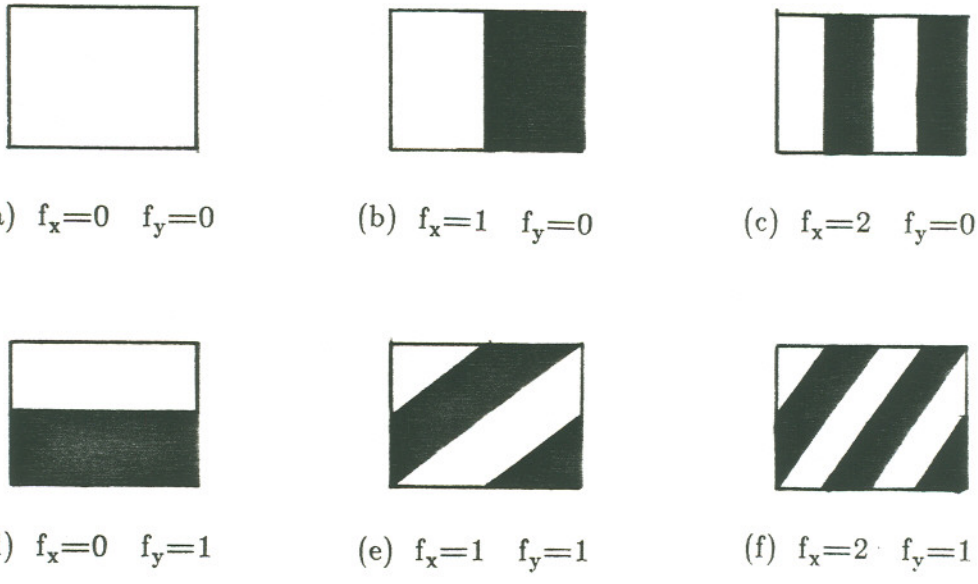


Fig.2.1 The concepts of horizontal and vertical frequency with simple stationary images

$f_d = f_x + f_y$, but in the other direction it is zero. Therefore, when we talk about the diagonal frequency, the diagonal direction should be specified. Usually, the highest diagonal frequency is a function of vertical frequency and horizontal frequency.

2.2. Video Signal Spectra

To describe video signal, three kinds of spectra are often used: one dimensional spectrum, two dimensional spectrum and three dimensional spectrum.

2.2.1. One Dimensional Spectrum

Since the picture is composed of scanning lines, the video signal can be considered as a serial signal along a long one dimensional line (see Fig.2.2). Considering it as a one dimensional signal, the spectrum can also be described in one dimension. For instance, if the transmitted picture is a stationary scene, the video waveform will repeat itself every frame, then the spectrum must be discrete and consist of harmonics of f_{frame}. For interlaced systems, blanking can cause field repetition so that its spectrum must have harmonics of $f_{field} = 2\ f_{frame}$. If the scene consists of constant vertical intensity, then every line has the same pattern and the spectrum of it will consist of line in frequency domain at the interval of f_{line}. For the typical stationary scene, its frequency spectrum will look like Fig.2.3(a). If the transmitted picture is changing, other frequencies will be introduced and the spectrum will become less discrete as shown in Fig.2.3(b).

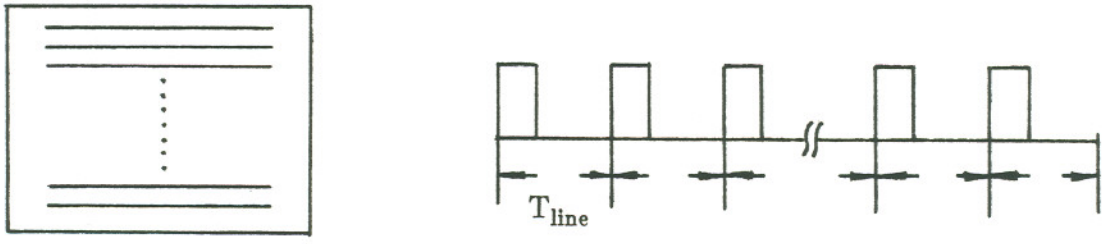
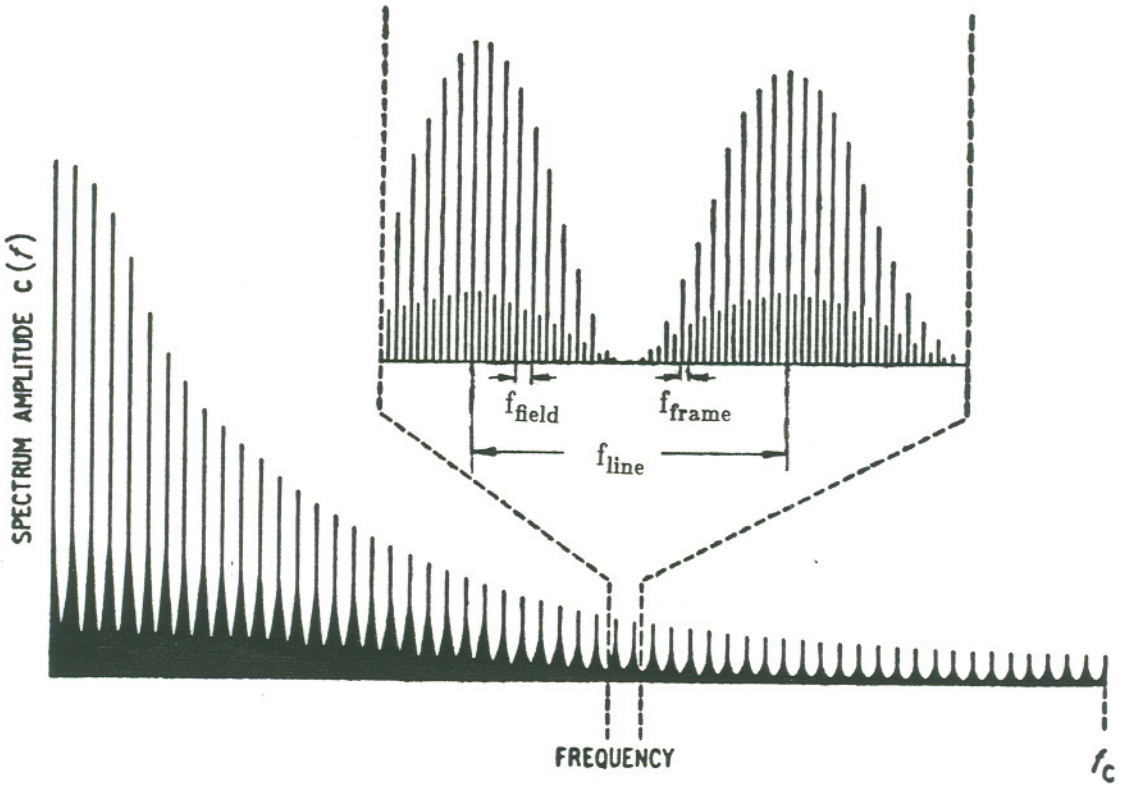
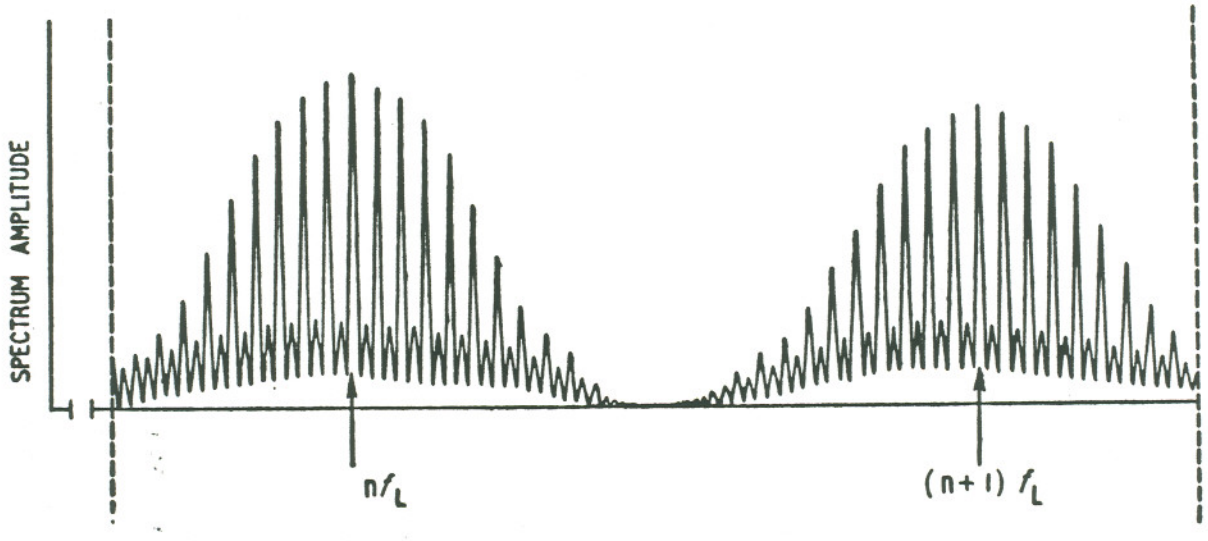


Fig.2.2 The video signal in one dimension



(a) NTSC frequency spectrum for a typical stationary scene



(b) NTSC frequency spectrum for a picture which is changing

Fig.2.3 NTSC frequency spectrum

2.2.2. Two Dimensional Spectrum

The one dimensional view of a video signal is not sufficient to analyze the scene that it represents. The two dimensional spectrum is often used to describe spatial patterns of the scene. For a fixed spatial pattern denoted i(x,y), two dimensional spectrum is represented as $I(f_x,f_y)$ in the frequency domain. It can also be considered as a instantaneous spectrum. The following briefly describes several two dimensional spectra of specific patterns which commonly occur in television pictures, such as edges, lines, textures and periodic structures.

2.2.2.1. Spectrum of Edges

The simplest pattern in a scene is a step edge (see Fig.2.4(a)). It can be described as

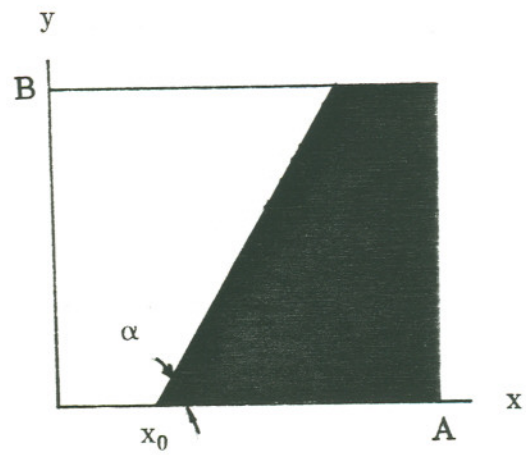
$$i(x,y) = \begin{cases} 0 & \text{when} \quad x < y\ \mathrm{ctg}(\alpha) + x_0 \\ 1 & \text{when} \quad x \geq y\ \mathrm{ctg}(\alpha) + x_0 \end{cases} \tag{2.2}$$

In order to easily deal with the problem, we can express the intensity as

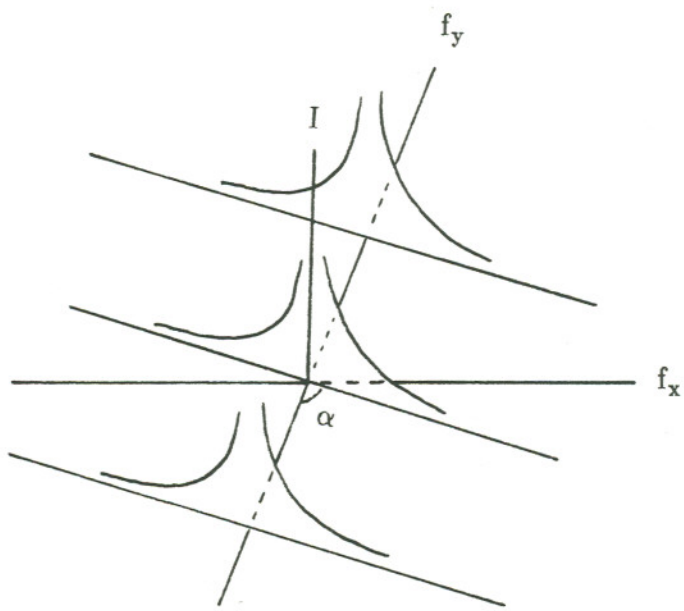
$$i(x,y) = \begin{cases} 0 & \text{when} \quad x < y\ \mathrm{ctg}(\alpha) + x_0 \\ e^{-\beta x} & \text{when} \quad x \geq y\ \mathrm{ctg}(\alpha) + x_0 \text{ and } \beta \to 0 \end{cases} \tag{2.3}$$

where β approaches zero very slowly.

The spectrum in two dimensional frequency domain will be



(a) Signal of an edge



(b) Spectrum of an edge

Fig.2.4 Singal and spectrum of an edge

$$I(f_x,f_y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} i(x,y)\, e^{-j2\pi(xf_x + yf_y)}\, dxdy$$

$$= \int_{-\infty}^{\infty} \lim_{B\to\infty} \int_{y\ ctg(\alpha)+x_0}^{B} \lim_{\beta\to 0} e^{-\beta x}\, e^{-j2\pi(xf_x+yf_y)}\, dxdy$$

$$= \int_{-\infty}^{\infty} \lim_{B\to\infty} \lim_{\beta\to 0} \int_{y\ ctg(\alpha)+x_0}^{B} e^{-\beta x}\, e^{-j2\pi(xf_x+yf_y)}\, dxdy$$

$$= \int_{-\infty}^{\infty} \lim_{B\to\infty} \lim_{\beta\to 0} \frac{e^{(-\beta-j2\pi f_x)B} - e^{(-\beta-j2\pi f_x)(y\ ctg(\alpha)+x_0)}}{-\beta-j2\pi f_x} e^{-j2\pi y f_y} dy$$

$$= \int_{-\infty}^{\infty} \frac{e^{-j2\pi f_x(y\ ctg(\alpha)\ +\ x_0)}}{-j2\pi f_x} e^{-j2\pi y f_y} dy \quad \text{when } \beta\to 0 \text{ with a slower rate}$$

$$= \frac{e^{-j2\pi f_x x_0}}{j2\pi f_x}\delta(f_y + f_x ctg(\alpha)) \tag{2.4}$$

where B is boundary along vertical direction.

The magnitude of the spectrum is concentrated on a line $f_y + f_x ctg\alpha = 0$ with its value proportional to $\frac{1}{f_x}$. If the picture is made up of scanning lines, as is the case for TV pictures, then it can be considered to be sampled in vertical direction with the sampling rate of $1/f_{line}$, in which f_{line} is the line scanning frequency. Therefore, in frequency domain, the spectrum becomes periodic one with the space nf_{line} (n is integer number) (see Fig.2.4(b)) as follows

$$I(f_x,f_y) = \frac{e^{-j2\pi f_x x_0}}{j2\pi f_x}\, \delta(f_y + f_x ctg(\alpha) + nf_{line}) \tag{2.5}$$

In practice, the boundary of integral is limited so that the spectrum broadens around line. The spectrum is given below.

$$I(f_x,f_y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} i(x,y)\, e^{-j2\pi(xf_x+yf_y)}\, dxdy$$

$$= \int_{0}^{B}\int_{y\ ctg(\alpha)+x_0}^{A} e^{-j2\pi(xf_x+yf_y)}\, dxdy$$

$$= \int_{0}^{B} \frac{e^{-j2\pi Af_x} - e^{-j2\pi f_x(y\ ctg(\alpha)+x_0)}}{-j2\pi f_x}\, e^{-j2\pi yf_y}\, dy$$

$$= \frac{e^{-j2\pi Af_x}[e^{-j2\pi Bf_y} - 1]}{(-j2\pi f_x)(-j2\pi f_y)} - \frac{e^{-j2\pi f_x x_0}[e^{-j2\pi B(f_y+f_x ctg(\alpha))} - 1]}{(-j2\pi f_x)[-j2\pi(f_y + f_x ctg(\alpha))]} \quad (2.6a)$$

It can also be expressed as a convolution of $I_\infty(f_x,f_y)$ with infinity boundaries (Eq.(2.5)) and a sinc function corresponding to limited boundaries A and B.

$$I(f_x,f_y) = \frac{1}{2\pi} I_\infty(f_x,f_y) * [AB\ Sa(\pi f_x A)\ Sa(\pi f_y B)\ e^{-j\pi Af_x}\ e^{-j\pi Bf_y}] \quad (2.6b)$$

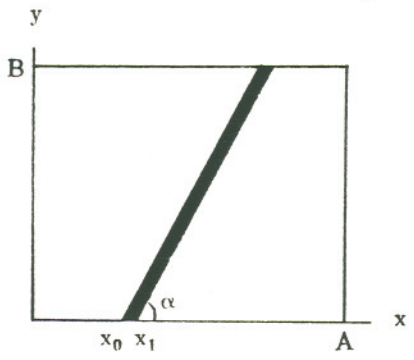
where

$$Sa(x) \equiv \frac{\sin x}{x} \quad \text{(sinc function)} \quad (2.7)$$

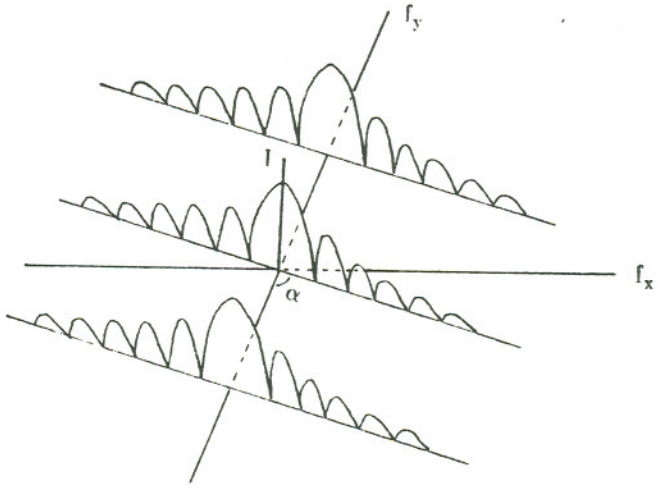
2.2.2.2. Spectrum of Lines

One of commonly used test signals is a strips pattern in two dimensions. Let us look at the spectrum of a simple example --- one line with a certain angle. Suppose the picture is a strip as in Fig.2.5(a), then its two dimensional Fourier

(a) Signal of a line



(b) Spectrum of a line



(c) Spectrum of a line with boundaries

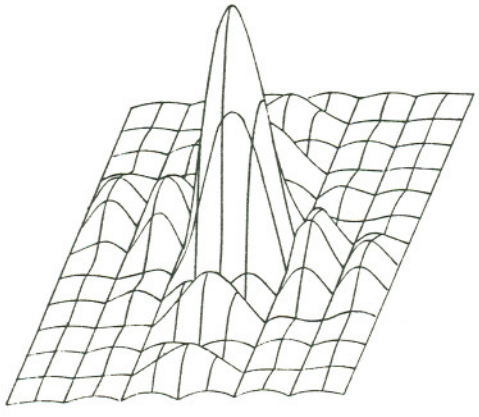


Fig.2.5 Signal and spectrum of a line

transform will be

$$I(f_x,f_y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} i(x,y)\, e^{-j2\pi(xf_x+yf_y)}\, dxdy$$

$$= \int_{-\infty}^{\infty}\int_{y\,ctg(\alpha)+x_0}^{y\,ctg(\alpha)+x_1} e^{-j2\pi(xf_x+yf_y)}\, dxdy$$

$$= \int_{-\infty}^{\infty} \frac{e^{-j2\pi f_x(y\,ctg(\alpha)+x_1)} - e^{-j2\pi f_x(y\,ctg(\alpha)+x_0)}}{-j2\pi f_x} e^{-j2\pi yf_y}\, dy$$

$$= \frac{e^{-j2\pi f_x x_1} - e^{-j2\pi f_x x_0}}{-j2\pi f_x}\, \delta(f_y + f_x ctg(\alpha))$$

$$= e^{-j\pi f_x(x_1-x_0)}\,(x_1 - x_0)\, Sa[\pi f_x(x_1 - x_0)]\, \delta(f_y + f_x ctg(\alpha)) \qquad (2.8)$$

So, the spectrum is still on the line $f_y + f_x ctg(\alpha) = 0$ but in the form of sinc function. Obviously, the spectrum is concentrated on a line $f_y = -f_x ctg(\alpha)$ with the magnitude $(x_1-x_0)Sa[\pi f_x(x_1-x_0)]$ as shown in Fig.2.5(b).

Again, the scanning lines cause this spectrum to repeat with the period of f_{line}, so that

$$I(f_x,f_y) = e^{-j\pi f_x(x_1-x_0)}(x_1-x_0)Sa[\pi f_x(x_1-x_0)]\delta(f_y+f_x ctg(\alpha)+nf_{line}) \qquad (2.9)$$

When the boundary is confined, the spectrum is modified as shown below and plotted in Fig.2.5(c).

$$I(f_x,f_y) = \int_{0}^{B}\int_{y\,ctg(\alpha)+x_0}^{y\,ctg(\alpha)+x_1} e^{-j2\pi(xf_x+yf_y)}\, dxdy$$

$$= \int_0^B \frac{e^{-j2\pi f_x(y\ ctg(\alpha)+x_1)} - e^{-j2\pi f_x(y\ ctg(\alpha)+x_0)}}{-j2\pi f_x} e^{-j2\pi y f_y}\, dy$$

$$= \frac{e^{-j2\pi f_x x_1} - e^{-j2\pi f_x x_0}}{-j2\pi f_x} \frac{e^{-j2\pi B(f_y+f_x ctg(\alpha))} - 1}{-j2\pi(f_y + f_x ctg(\alpha))}$$

$$= e^{-j\pi f_x(x_1-x_0)}\, e^{-j\pi B(f_y+f_x ctg(\alpha))}$$

$$\times (x_1-x_0)\, Sa[\pi f_x(x_1-x_0)]\, B\, Sa[\pi(f_y+f_x ctg(\alpha))B] \qquad (2.10)$$

2.2.2.3. Spectrum of A Rectangle

For the rectangular patterns (Fig.2.6(a)), the spectrum is given by

$$I(f_x,f_y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} i(x,y)\, e^{-j2\pi(xf_x+yf_y)}\, dxdy$$

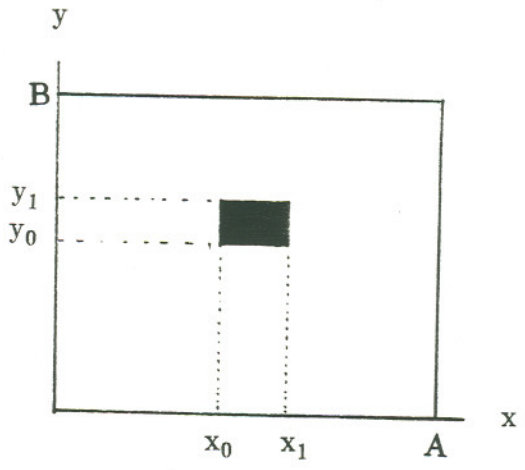
$$= \int_{y_0}^{y_1}\int_{x_0}^{x_1} e^{-j2\pi(xf_x+yf_y)}\, dxdy$$

$$= \frac{e^{-j2\pi f_x x_1} - e^{-j2\pi f_x x_0}}{-j2\pi f_x} \frac{e^{-j2\pi f_y y_1} - e^{-j2\pi f_y y_0}}{-j2\pi f_y}$$

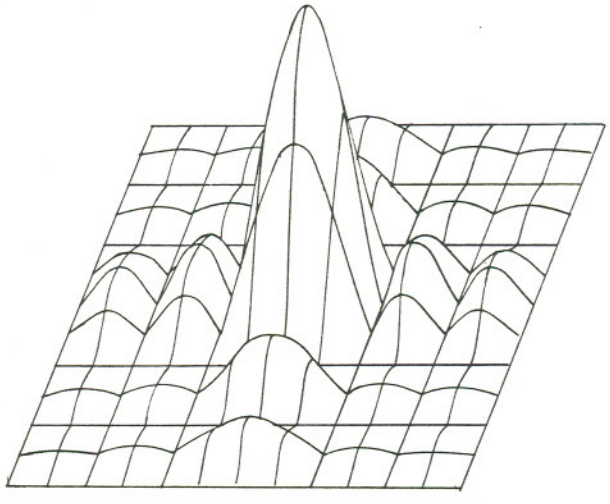
$$= e^{-j\pi[f_x(x_1-x_0)+f_y(y_1-y_0)]}$$

$$\times (x_1-x_0)\,(y_1-y_0)\, Sa[\pi f_x(x_1-x_0)]\, Sa[\pi f_y(y_1-y_0)] \qquad (2.11)$$

Sinc function makes the spectrum in ripple form (Fig.2.6).



(a) Signal of a rectangle



(b) Spectrum of a rectangle

Fig.2.6 Spectrum of a rectangle

2.2.2.4. Spectrum of Texture

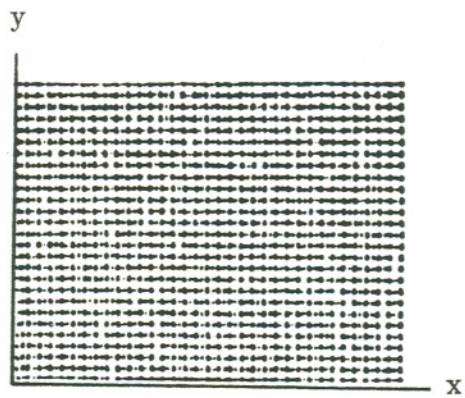
The texture patterns can be considered as a summation of rectangles. In a regular texture, the periodic rectangles result the spectrum sampled in frequency domain. As the texture is limited within boundaries, its spectrum is widened. The spectrum is hill-like. Sometimes, the texture pattern can be described as a sampling of rectangular patterns. Via zero padding, its spectrum will be a hill-like one in which the hill is located at harmonics of sampling frequency (Fig.2.7). If the pattern has pseudorandom textures, its spectrum still appears continuous and hill-like. If the texture has a certain directionality, its spectrum will also be oriented in the corresponding direction in the frequency domain.

2.2.2.5. Zone Plate (ZP)

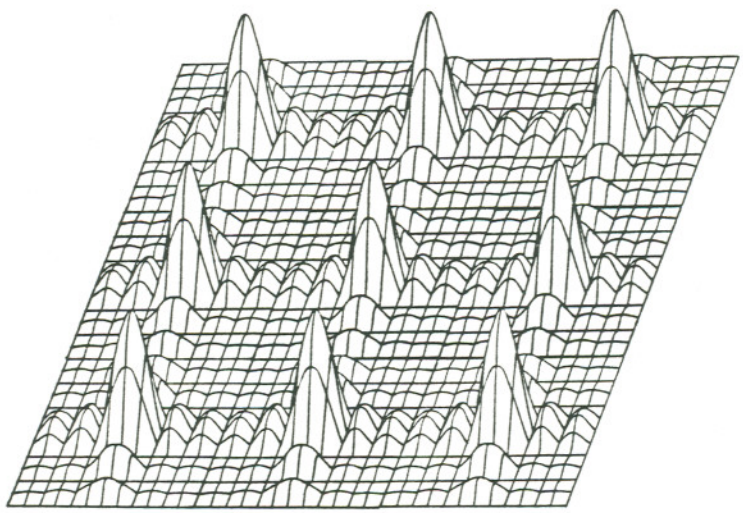
The zone plate[3] pattern is a two dimensional linear sweep of spatial frequency. The reason that it is so attractive is that it can be taken as a test method for a quick appraisal of system's frequency characteristics since the coordinates x and y also happen to be the two horizontal and vertical frequencies (as explained later). The typical patterns are circular zone plate (CZP) and hyperbolic zone plate (HZP) whose intensities are generated by Eq.(2.12) and Eq.(2.13) respectively.

$$\text{Circular: } i(x,y) = \cos[\pi(x^2 + y^2)] \tag{2.12}$$

where the center of the pattern is at the origin (0,0) as shown in Fig.2.8.



(a) Signal of a texture



(b) Spectrum of a texture

Fig.2.7 Spectrum of a texture

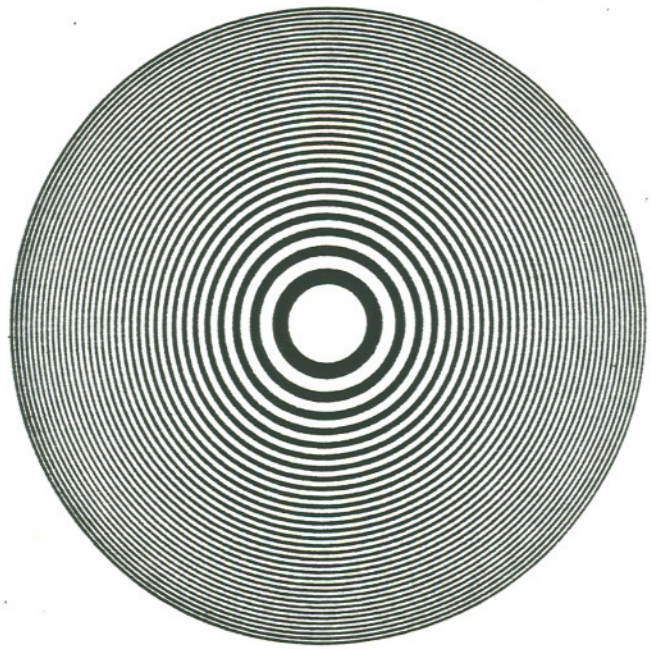


Fig.2.8 A circular zone plate

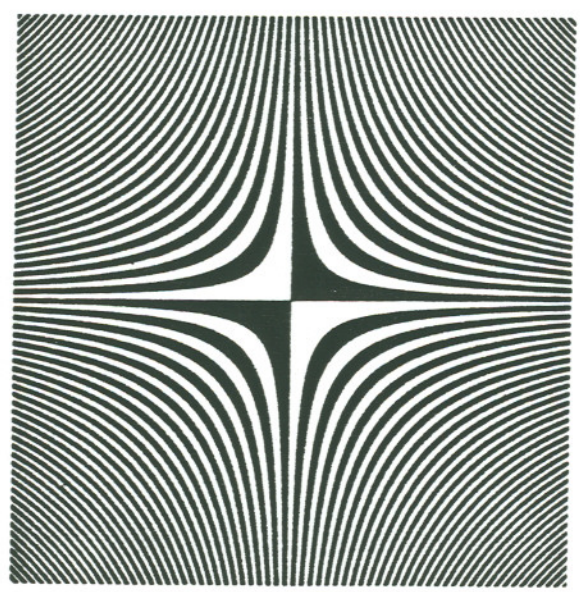


Fig.2.9 A hyperbolic zone plate

$$\text{Hyperbolic: } i(x,y) = \cos(2\pi xy) \tag{2.13}$$

where the center of the pattern is at the origin (0,0) as shown in Fig.2.9.

Since the horizontal and vertical frequency components f_x and f_y can be solved as partial derivatives of the phase with respect to x and y, f_x and f_y are obtained by

$$\text{Circular: } f_{xc} = \frac{1}{2\pi}\frac{\partial}{\partial x}[\pi(x^2 + y^2)] = x \tag{2.14a}$$

$$f_{yc} = \frac{1}{2\pi}\frac{\partial}{\partial y}[\pi(x^2 + y^2)] = y \tag{2.14b}$$

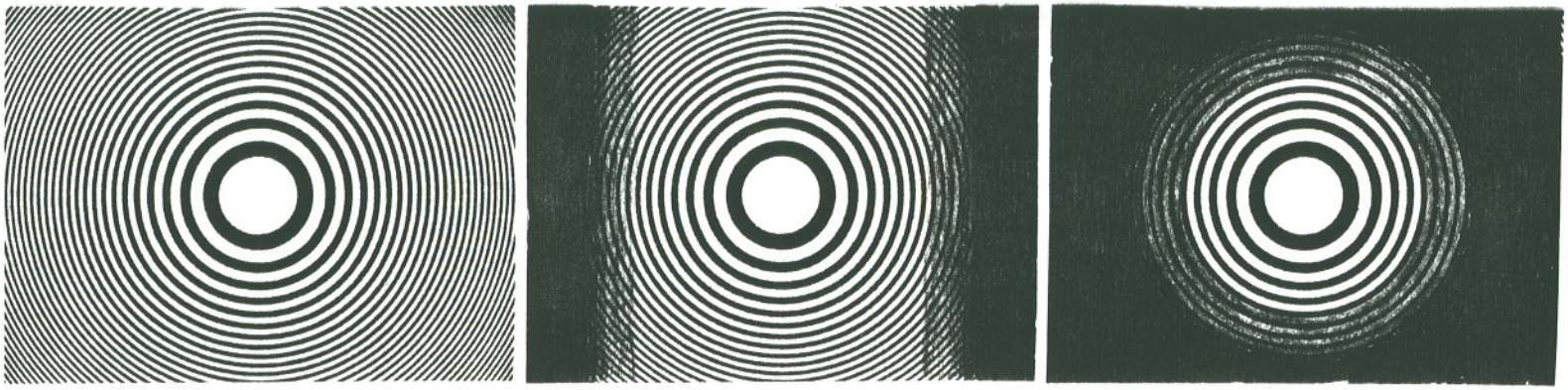
$$\text{Hyperbolic: } f_{xh} = \frac{1}{2\pi}\frac{\partial}{\partial x}[2\pi xy] = y \tag{2.15a}$$

$$f_{yh} = \frac{1}{2\pi}\frac{\partial}{\partial y}[2\pi xy] = x \tag{2.15b}$$

The spectrum of it is

$$I(f_x,f_y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} i(x,y)e^{-j2\pi(f_x x + f_y y)}dxdy \tag{2.16}$$

This characteristics results displaying the spatial frequency characteristics directly on a cathode-ray tube (CRT). Take the circular zone plate as an example (see Fig.2.10). Fig.2.10(a) is an original circular zone plate. If a horizontal low pass filter is applied, the effect of the filter (see Fig.2.10(b)) is to pass all frequencies between the parallel vertical lines and beyond the pass region (at high



(a) Origin (b) After horizontal low pass filter (c) After circuler low pass filter

Fig.2.10 The effect of zone plate under different frequencies

absolute horizontal coordinates) the display changes to a uniform gray. The transition band results the edges of those vertical lines blurred. If a low pass filter is circular symmetry, the image of CZP becomes circular shadow (see Fig.2.10(c)).

2.2.3. Three Dimensional Spectrum

To describe the changing scene, the three dimensional space x, y and t is used. Let $i(x,y,t)$ represent an intensity distribution and $I(f_x,f_y,f_t)$ as its Fourier transform. For the sake of four parameters f_x,f_y,f_t and I, four dimensional space is required to represent the spectrum. However, the visible space is only of three dimensions so that the spectrum is traced by fixing one parameter and displaying the other three. For instance, at a given time $t=t_0$, the picture is represented by $i(x,y,t_0)$ and $I(f_x,f_y)\,|_{t=t_0}$; or fixing a column ($x=x_0$), to see what is going on in vertical-temporal direction; or just see coarsely what frequency components might occur in f_x,f_y and f_t domain. Now let us consider a few simple examples.

2.2.3.1. Spectrum of Cube

Imagine that we shoot a blackboard on a wall successively. What happens is that we get the same picture at different times. These pictures represent a cubic shaped scene in three dimensions. Its expression is given by

$$i(x,y,t) = \begin{cases} 1 & \text{when } x_0 \le x \le x_1, y_0 \le y \le y_1 \text{ and } t_0 \le t \le t_1 \\ 0 & \text{otherwise} \end{cases} \tag{2.17}$$

and its Fourier transform is

$$I(f_x,f_y,f_t) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} i(x,y,t)\, e^{-j2\pi(xf_x+yf_y+tf_t)}\, dxdydt$$

$$= \int_{t_0}^{t_1}\int_{y_0}^{y_1}\int_{x_0}^{x_1} e^{-j2\pi(xf_x+yf_y+tf_t)}\, dxdydt$$

$$= \frac{e^{-j2\pi f_x x_1} - e^{-j2\pi f_x x_0}}{-j2\pi f_x} \frac{e^{-j2\pi f_y y_1} - e^{-j2\pi f_y y_0}}{-j2\pi f_y} \frac{e^{-j2\pi f_t t_1} - e^{-j2\pi f_t t_0}}{-j2\pi f_t}$$

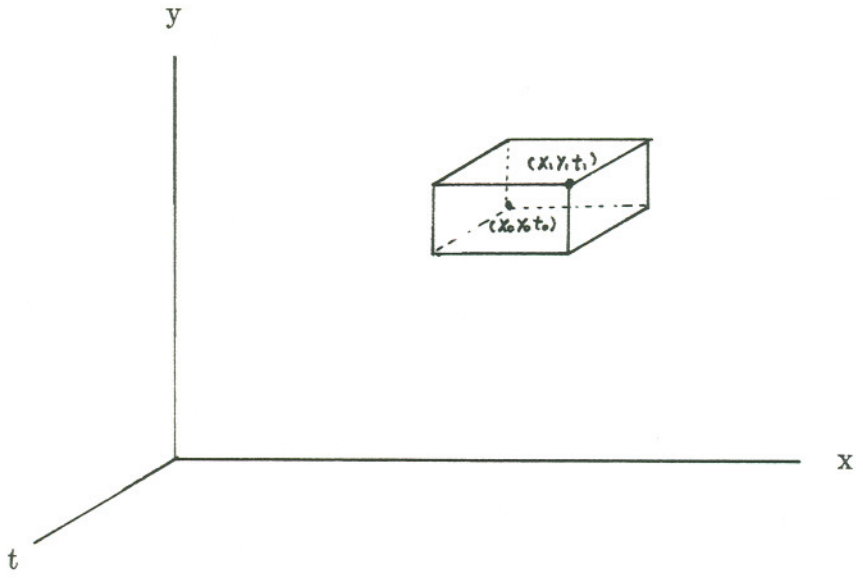
$$= e^{-j\pi[f_x(x_1-x_0)+f_y(y_1-y_0)+f_t(t_1-t_0)]}\,(x_1 - x_0)\,(y_1 - y_0)\,(t_1 - t_0)$$

$$\times \mathrm{Sa}[\pi f_x(x_1 - x_0)]\ \mathrm{Sa}[\pi f_y(y_1 - y_0)]\ \mathrm{Sa}[\pi f_t(t_1 - t_0)] \qquad (2.18)$$

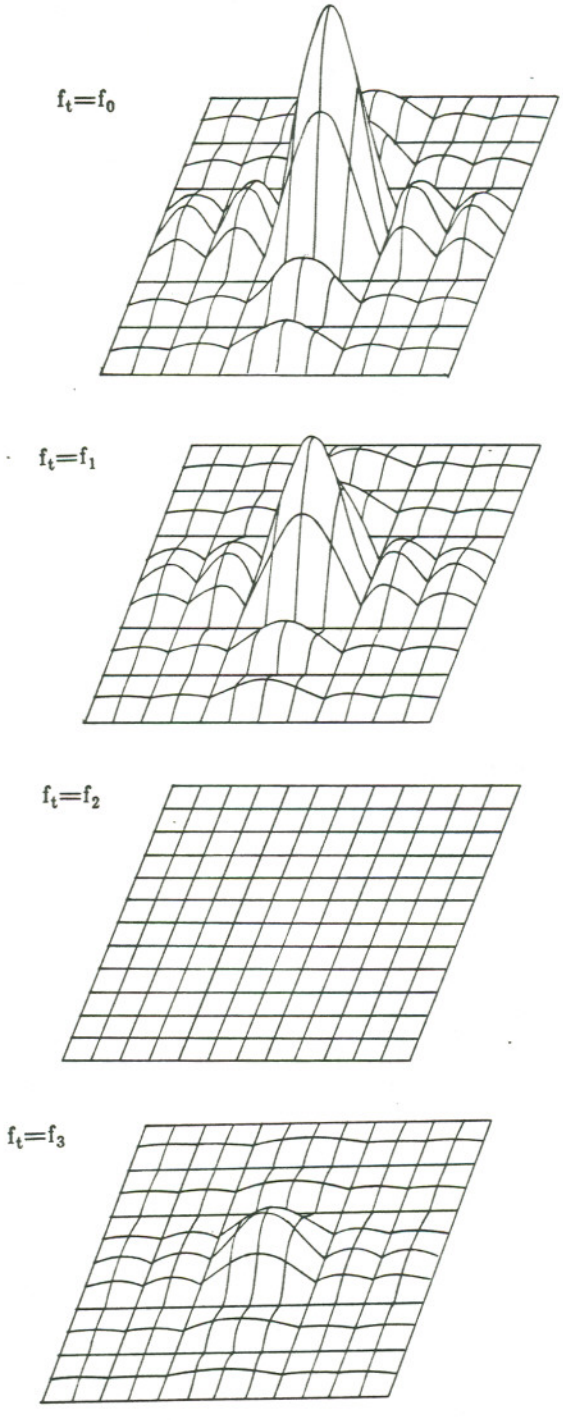
Think of the spectrum described by frequency domain (f_x,f_y,f_t). If we chose the method of tracing the spectrum described by (I,f_x,f_y) for a given f_t, the spectra will look the same except for a different gain (see Fig.2.11(d)).

2.2.3.2. To and Fro Zone Plate (TFZP)

As is well known, ZP is one method to display two dimensional frequency characteristics. The three dimensional frequency characteristics can be observed by simply applying horizontal or vertical to-and-fro motion to the zone plate[4]. Consider the circular zone plate as an example.



(a) Signal of a cube



(b) Spectrum of a cube in f_x , f_y and I domain

Fig.2.11 Signal and spectrum of a cube

Suppose the motion is linear, an image can be generated as

$$f(x,y,t) = i(x-v_x t\ ,\ y-v_y t) = \cos[\pi((x-v_x t)^2 + (y-v_y t)^2)] \tag{2.19}$$

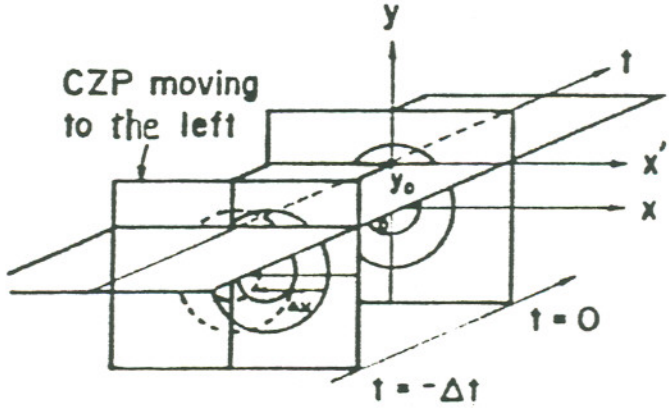
where v_x and v_y are the velocities in horizontal and vertical directions, respectively. Its three dimensional frequency spectrum is obtained by

$$\begin{aligned} F(f_x,f_y,f_t) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y,t)\, e^{-j2\pi(f_x x+f_y y+f_t t)}\, dxdydt \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} i(x-v_x t, y-v_y t, t)\, e^{-j2\pi(f_x x+f_y y+f_t t)}\, dxdydt \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} i(x,y,t)\, e^{-j2\pi(f_x x+f_y y)}\, e^{-j2\pi(v_x f_x+v_y f_y+f_t)t}\, dxdydt \\ &= I(f_x,f_y)\, \delta(v_x f_x+v_y f_y+f_t) \end{aligned} \tag{2.20}$$

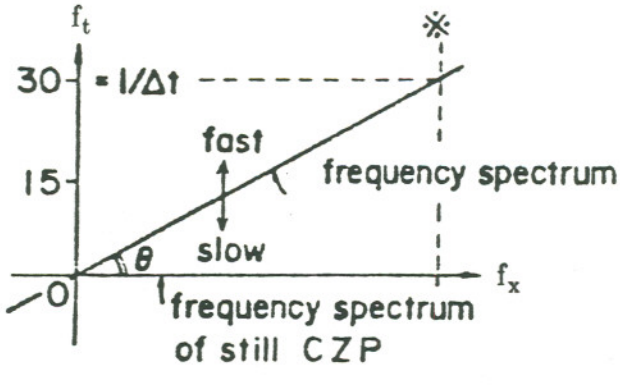
Therefore, the three dimensional frequency spectrum exists only on the plane $v_x f_x+v_y f_y+f_t=0$ in the three dimensional frequency domain.

Let us illustrate it by an example. Consider the horizontal movement of a circular zone plate toward the left with a speed v (see Fig.2.12(a)). Because the circular zone plate is displaced linearly by a translation movement, the temporal frequency f_t is proportional to the horizontal frequency f_x (see Fig.2.12(b)) as

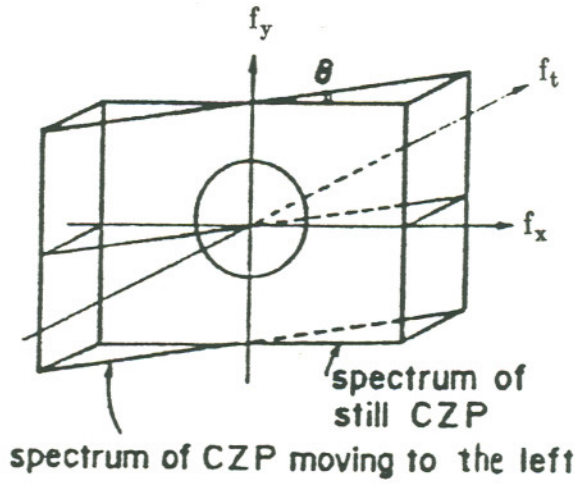
$$f_t = \frac{1}{\Delta t} = \frac{1}{\frac{\Delta x}{v}} = v f_x \equiv \mathrm{tg}\theta\ f_x \tag{2.21}$$



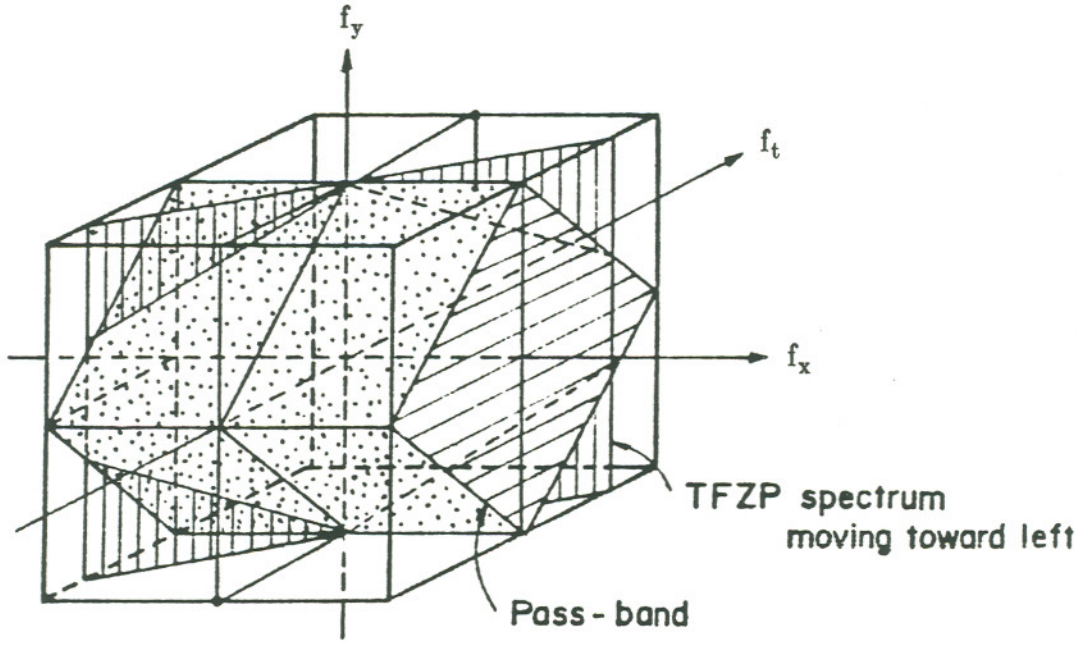
(a) TFZP scheme



(b) Frequency spectrum of the TFZP in a 2-D frequency domain



(c) Frequency spectrum of the TFZP in a 3-D frequency domain



(d) Effect of a 3-D filter in a 3-D frequency domain

Fig.2.12 Scheme and spectrum of to-and-fro zone plate

The faster the movement, the higher the f_t goes and the greater the angle is. Fig.2.12(c) gives the relation between the stationary zone plate and the to-and-fro moving zone plate. If the angle is altered from 0 to 180°, the continuous zone plate spectra consist of a three dimensional spectrum. The effect of a three dimensional horizontal-vertical-temporal filter in a three dimensional frequency domain is shown as an example (in Fig.2.12(d)).

2.2.4. Effects of Sampling

TV signal is inherently a sampled signal in the vertical and temporal direction. Many times it is also sampled in the horizontal direction. Therefore, it is essential to understand the effects of the various sampling structures generally in the frequency domain. One of the results of the sampling is periodic repetition of the spectrum. According to the repeat position in three dimensional spectrum domain, a feasible optional three dimensional spectral domain can be constructed so as to avoid aliasing. Here, we give some examples with regard to sampling structures and their repeat positions in frequency domain in which normalization will be used as follows: the distance between adjacent points in the horizontal direction, the distance of two consecutive lines as well as the duration of one field are equal to unit 1, or $x=1$, $y=1$ and $t=1$. The sampling frequencies corresponding to these unit values are unit 1, or $f_x=1$, $f_y=1$ and $f_t=1$, respectively.

2.2.4.1. Sampling in Two-Dimension

Assume that a picture shown in Fig.2.13(a)[5] consists of n×m pixels, where n scanning lines can support frequencies up to n/2 c/ph and m pixels in row can support frequencies up to m/2 c/pw. Consider one interlaced field, i.e. sampling in vertical direction with half rate of progressive scan field. Since sampling causes spectrum repetition, the repeat point will be located on the harmonics of the sampling frequency. The highest frequency in the spectrum without aliasing will decrease to half of the original one given by Fig.2.13(b). So does the sampling in the horizontal direction. One interesting method of sampling is the diagonal sampling called the quincunx sampling (Fig.2.13(c)). The high diagonal frequencies are lost but both high horizontal and high vertical frequencies are preserved. If the quincunx subsampling is applied, the data rate gets reduced by a factor of 2 (Fig.2.13(d)). This is a commonly used approach to compress the TV signal.

2.2.4.2. Sampling in Three-Dimension

Let us start by considering the video information to be confined to a rectangular box called a lattice of dimensions A, B and C in x, y and t space. The complex exponential form of the Fourier Series gives the harmonics f_x, f_y and f_t which are located on a rectangular grid called the reciprocal lattice[6,7] as the reciprocal relation between wavelength and frequency. The grids are spaced $\frac{1}{A}$, $\frac{1}{B}$ and $\frac{1}{C}$ in the three dimensional frequency domain as shown in Fig.2.14.

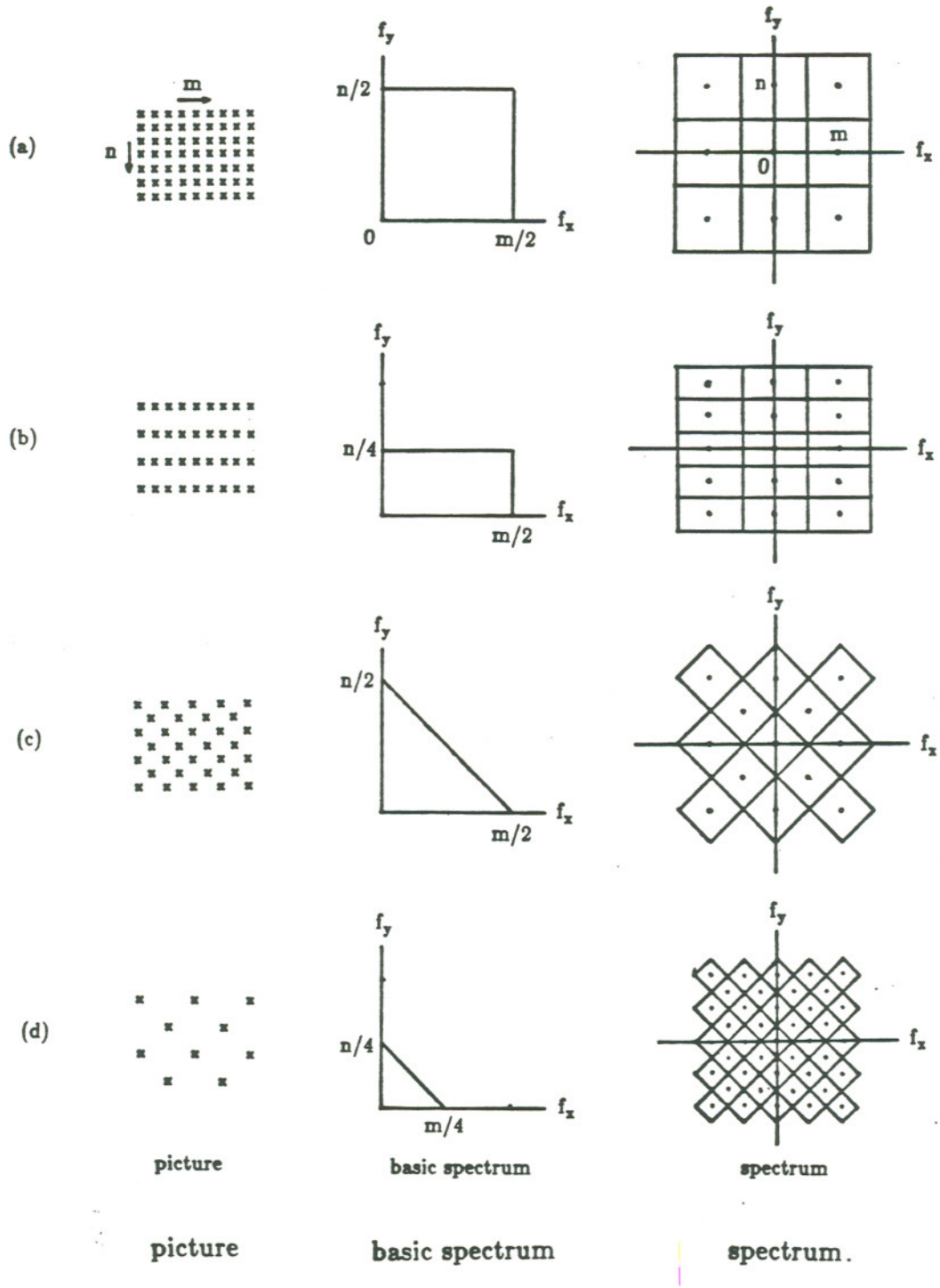


Fig.2.13 Signal and spectrum of sampling in two dimension

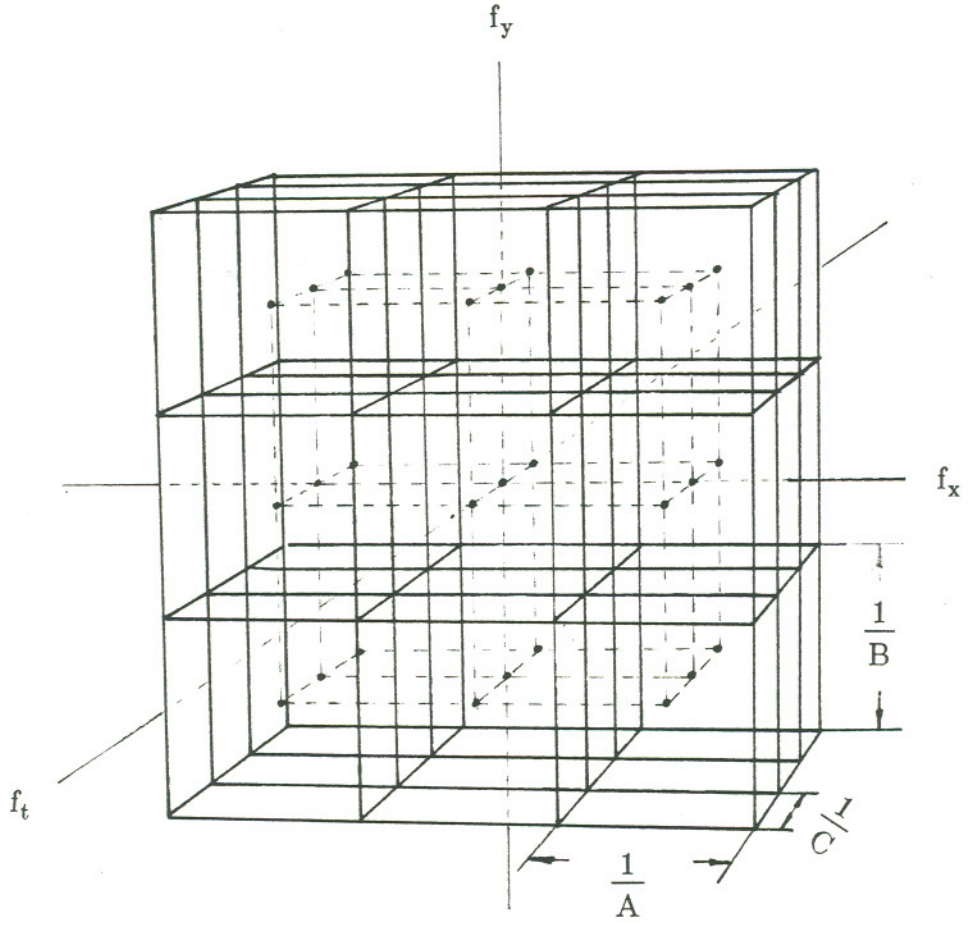


Fig.2.14 Frequency reciprocal lattice

The box $\frac{1}{A}$, $\frac{1}{B}$ and $\frac{1}{C}$ is centered at repetition points.

By means of the mathematic tools, we can find out the repeat positions accurately. At the beginning, assume

$|\underline{T}|$ is the determinant of matrix $\underline{T}$

$\underline{T}^t$ is the transpose matrix

$\underline{T}^{-1}$ is the inverse matrix

$\underline{T}^{-t}$ is the transpose of the inverse matrix

$\underline{I}$ is the unit matrix

Before considering a signal $s(\underline{x})$ whose Fourier transform is $S(\underline{f})$ where $\underline{x}=(x,y,t)^t$ and $\underline{f}=(f_x,f_y,f_t)^t$, let us look at a periodic sampling impulse signal $\phi(\underline{x})$ given by

$$\phi(\underline{x}) = \sum_{\underline{k}} \gamma_{\underline{k}} \delta(\underline{x}-\underline{L}\underline{k}) \tag{2.22}$$

where $\gamma_{\underline{k}} = \gamma_{\underline{b}+\underline{T}\underline{i}} = \gamma_{\underline{b}}$ which is the periodic weight of Dirac pulses with basic weight $\gamma_{\underline{b}}$ in one period. The simplest sampling impulse signal is Dirac pulses with unit weight.

$\underline{k}$, $\underline{i}$ and $\underline{b}$ are integer vectors.

$\underline{L}$ is sampling space scale matrix and is assumed to be an identical matrix in the example.

$\underline{T}$ is the translation scale matrix and it is a diagonal matrix.

Then

$$\phi(\underline{x}) = \sum_{\underline{k}} \gamma_{\underline{k}}\, \delta(\underline{x}-\underline{L}\underline{k})$$

$$= \sum_{\underline{l}} \sum_{\underline{b}} \gamma_{\underline{b}}\, \delta[\underline{x}-\underline{L}(\underline{b}+\underline{T}\underline{l})]$$

$$= \sum_{\underline{l}} \sum_{\underline{b}} \gamma_{\underline{b}}\, \delta(\underline{x}-\underline{L}\underline{b}-\underline{L}\underline{T}\underline{l})$$

$$= \sum_{\underline{l}} \phi_{\underline{b}}(\underline{x}-\underline{L}\underline{T}\underline{l}) \qquad (2.23)$$

where

$$\phi_{\underline{b}}(\underline{x}) \equiv \sum_{\underline{b}} \gamma_{\underline{b}}\, \delta(\underline{x}-\underline{L}\underline{b}) \qquad (2.24)$$

We take the commonly used sampling --- the line interlaced sampling as an example (Fig.2.15(b1) has shown its patterns of sampling lattice and reciprocal lattice).

Assume $\underline{L} = \underline{I}$

Choose

$$\underline{T} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

$$\gamma_{\underline{b}} = \begin{cases} 1 & \text{when } \underline{b} = [\,0\;0\;0\,]^t \text{ and } \underline{b} = [\,0\;1\;1\,]^t \\ 0 & \text{otherwise} \end{cases}$$

where $\underline{T}^{-1}\underline{b} \in \underline{I}$ since $\gamma_{\underline{b}}$ are basic weights in one period.

As $\gamma_{\underline{k}}$ is periodic,

$$\phi(\underline{x}) = \sum_{\underline{l}} \sum_{\underline{b}} \gamma_{\underline{b}}\, \delta(\underline{x} - \underline{Lb} - \underline{LTl})$$

$$= \sum_{\underline{l}} \left\{ \delta\left(\underline{x} - \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix} \underline{l} \right) + \delta\left(\underline{x} - \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix} \underline{l} \right) \right\}$$

Now, let us find the Fourier transform of $\phi(\underline{x})$. Since $\phi(\underline{x})$ is a periodic impulse signal, it can be expanded in a Fourier Series as

$$\phi(\underline{x}) = \sum_{\underline{l}} \phi_{\underline{b}}(\underline{x} - \underline{LTl})$$

$$\equiv |\underline{LT}|^{-1} \sum_{\underline{i}} \Gamma_{\underline{i}}\, e^{j2\pi \underline{i}^t (\underline{LT})^{-1} \underline{x}} \qquad (2.25)$$

where the coefficient is

$$\Gamma_{\underline{i}} = \int_{-\frac{\underline{LT}}{2}}^{\frac{\underline{LT}}{2}} \phi_{\underline{b}}(\underline{x})\, e^{-j2\pi \underline{i}^t (\underline{LT})^{-1} \underline{x}}\, d\underline{x}$$

$$= \int_{-\frac{\underline{LT}}{2}}^{\frac{\underline{LT}}{2}} \sum_{\underline{b}} \gamma_{\underline{b}}\, \delta(\underline{x} - \underline{Lb})\, e^{-j2\pi \underline{i}^t (\underline{LT})^{-1} \underline{x}}\, d\underline{x}$$

$$= \sum_{\underline{b}} \gamma_{\underline{b}} \int_{-\frac{\underline{LT}}{2}}^{\frac{\underline{LT}}{2}} \delta(\underline{x} - \underline{Lb})\, e^{-j2\pi \underline{i}^t (\underline{LT})^{-1} \underline{x}}\, d\underline{x}$$

$$= \sum_{\underline{b}} \gamma_{\underline{b}} \; e^{-j2\pi \underline{i}^t (\underline{LT})^{-1} \underline{Lb}}$$

$$= \sum_{\underline{b}} \gamma_{\underline{b}} \; e^{-j2\pi \underline{i}^t \underline{T}^{-1} \underline{b}} \qquad (\underline{T}^{-1}\underline{b} \in \underline{I}) \tag{2.26}$$

Apply the Fourier operator on $\phi(\underline{x})$ to derive its Fourier transform $\Phi(\underline{f})$ as

$$\Phi(\underline{f}) = \mathcal{F}\,[\phi(\underline{x})]$$

$$= \mathcal{F}\,[\,|\underline{LT}|^{-1} \sum_{\underline{i}} \Gamma_{\underline{i}} \, e^{j2\pi \underline{i}^t (\underline{LT})^{-1} \underline{x}}]$$

$$= |\underline{LT}|^{-1} \sum_{\underline{i}} \Gamma_{\underline{i}} F\,[\, e^{j2\pi \underline{i}^t (\underline{LT})^{-1} \underline{x}}]$$

$$= |\underline{LT}|^{-1} \sum_{\underline{i}} \Gamma_{\underline{i}} \, \delta(\underline{f} - [\underline{i}^t (\underline{LT})^{-1}]^t)$$

$$= |\underline{LT}|^{-1} \sum_{\underline{i}} \Gamma_{\underline{i}} \, \delta(\underline{f} - (\underline{LT})^{-t} \underline{i}) \tag{2.27}$$

As a sampled signal $s_s(\underline{x})$ can be constructed by multiplying original signal $s(\underline{x})$ and sampling impulses signal $\phi(\underline{x})$, the Fourier transform of sampled signal, $S_s(\underline{f})$ is equivalent to the convolution of $S(\underline{f})$ and $\Phi(\underline{f})$ in the frequency domain as

$$S_s(\underline{f}) = S(\underline{f}) * \Phi(\underline{f})$$

$$= S(\underline{f}) * [\,|\underline{LT}|^{-1} \sum_{\underline{i}} \Gamma_{\underline{i}} \, \delta(\underline{f} - (\underline{LT})^{-t} \underline{i})]$$

$$= |\underline{LT}|^{-1} \sum_{i} \Gamma_i \, S(\underline{f} - (\underline{LT})^{-t}\underline{i})] \tag{2.28}$$

Obviously, the spectrum of $S_s(\underline{x})$ is the repetition of $S(\underline{f})$ at the positions $(\underline{TL})^{-t}\underline{i}$. Based on these repeat positions, we can confine the signal so as to avoid aliasing. Let us get back to the above example to explain how to find the repeat positions. According to Eq.(2.26), the coefficients Γ_i are given by

$$\Gamma_{[0\,0\,0]^t} = 1 + e^{-j2\pi[0\,1\,1]\begin{bmatrix}1 & 0 & 0\\ 0 & 1/2 & 0\\ 0 & 0 & 1/2\end{bmatrix}\begin{bmatrix}0\\0\\0\end{bmatrix}} = 2$$

$$\Gamma_{[0\,0\,1]^t} = 1 + e^{-j2\pi[0\,1\,1]\begin{bmatrix}1 & 0 & 0\\ 0 & 1/2 & 0\\ 0 & 0 & 1/2\end{bmatrix}\begin{bmatrix}0\\0\\1\end{bmatrix}} = 0$$

$$\Gamma_{[0\,1\,0]^t} = 1 + e^{-j2\pi[0\,1\,1]\begin{bmatrix}1 & 0 & 0\\ 0 & 1/2 & 0\\ 0 & 0 & 1/2\end{bmatrix}\begin{bmatrix}0\\1\\0\end{bmatrix}} = 0$$

$$\Gamma_{[0\,1\,1]^t} = 1 + e^{-j2\pi[0\,1\,1]\begin{bmatrix}1 & 0 & 0\\ 0 & 1/2 & 0\\ 0 & 0 & 1/2\end{bmatrix}\begin{bmatrix}0\\1\\1\end{bmatrix}} = 2$$

$$\Gamma_{[1\,0\,0]^t} = 1 + e^{-j2\pi[0\,1\,1]\begin{bmatrix}1 & 0 & 0\\ 0 & 1/2 & 0\\ 0 & 0 & 1/2\end{bmatrix}\begin{bmatrix}1\\0\\0\end{bmatrix}} = 2$$

$$\Gamma_{[1\,0\,1]^t} = 1 + e^{-j2\pi[0\,1\,1]\begin{bmatrix}1 & 0 & 0\\ 0 & 1/2 & 0\\ 0 & 0 & 1/2\end{bmatrix}\begin{bmatrix}1\\0\\1\end{bmatrix}} = 0$$

$$\Gamma_{[1\,1\,0]^t} = 1 + e^{-j2\pi[0\,1\,1]\begin{bmatrix}1 & 0 & 0\\ 0 & 1/2 & 0\\ 0 & 0 & 1/2\end{bmatrix}\begin{bmatrix}1\\1\\0\end{bmatrix}} = 0$$

$$\Gamma_{[1\,1\,1]^t} = 1 + e^{-j2\pi[0\,1\,1]\begin{bmatrix}1 & 0 & 0\\ 0 & 1/2 & 0\\ 0 & 0 & 1/2\end{bmatrix}\begin{bmatrix}1\\1\\1\end{bmatrix}} = 2$$

Significant and effective coefficients are those $\Gamma_{\underline{i}}$ which are not equal to zero, i.e. $\Gamma_{[0\,0\,0]^t}$, $\Gamma_{[0\,1\,1]^t}$, $\Gamma_{[1\,0\,0]^t}$ and $\Gamma_{[1\,1\,1]^t}$. Corresponding to these non zero $\Gamma_{\underline{i}}$, we get the basic repeat positions in the form

$$P_{\underline{i}} \equiv (\underline{LT})^{-t}\underline{i} \tag{2.29}$$

i.e.

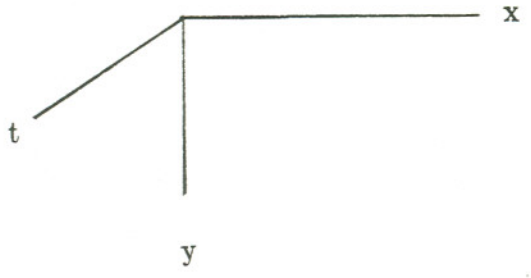
$$P_{[0\,0\,0]^t} = \begin{bmatrix}1 & 0 & 0\\ 0 & 1/2 & 0\\ 0 & 0 & 1/2\end{bmatrix}\begin{bmatrix}0\\0\\0\end{bmatrix} = \begin{bmatrix}0\\0\\0\end{bmatrix}$$

$$P_{[0\,1\,1]^t} = \begin{bmatrix}1 & 0 & 0\\ 0 & 1/2 & 0\\ 0 & 0 & 1/2\end{bmatrix}\begin{bmatrix}0\\1\\1\end{bmatrix} = \begin{bmatrix}0\\1/2\\1/2\end{bmatrix}$$

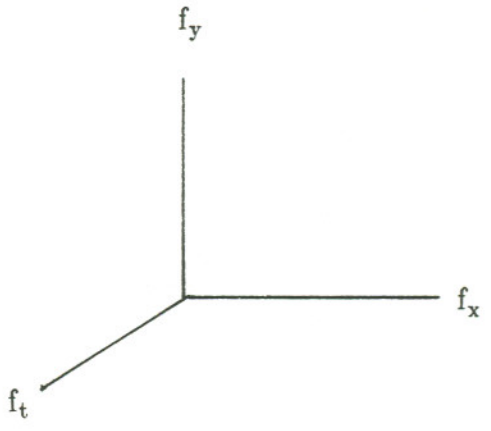
$$P_{[1\,0\,0]^t} = \begin{bmatrix}1 & 0 & 0\\ 0 & 1/2 & 0\\ 0 & 0 & 1/2\end{bmatrix}\begin{bmatrix}1\\0\\0\end{bmatrix} = \begin{bmatrix}1\\0\\0\end{bmatrix}$$

$$P_{[1\,1\,1]^t} = \begin{bmatrix}1 & 0 & 0\\ 0 & 1/2 & 0\\ 0 & 0 & 1/2\end{bmatrix}\begin{bmatrix}1\\1\\1\end{bmatrix} = \begin{bmatrix}1\\1/2\\1/2\end{bmatrix}$$

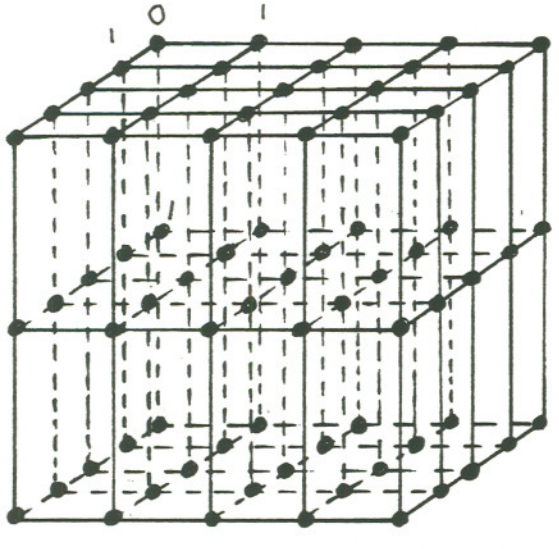
Since $\underline{T}^{-1}\underline{i}$ is confined in unit $\underline{I}$, the basic repeat points will be (0,0,0), (0,1/2,1/2) and unit corners. This means the repetition block in reciprocal space is composed of these basic repeat points with frequency translation matrix $\underline{I}$ (see Fig.2.15(b2)).

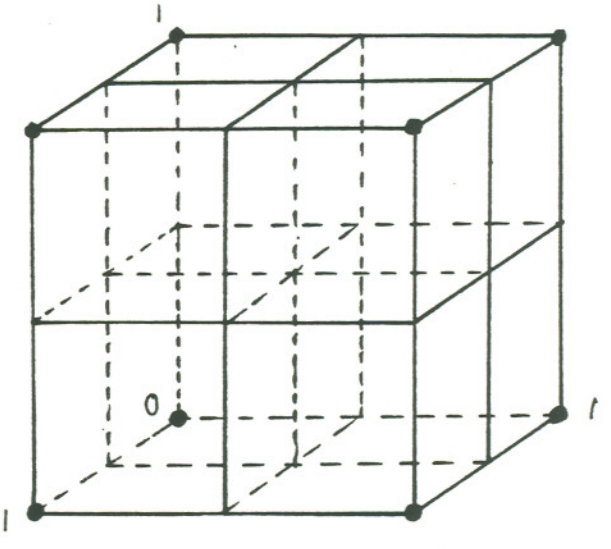


sampling pattern

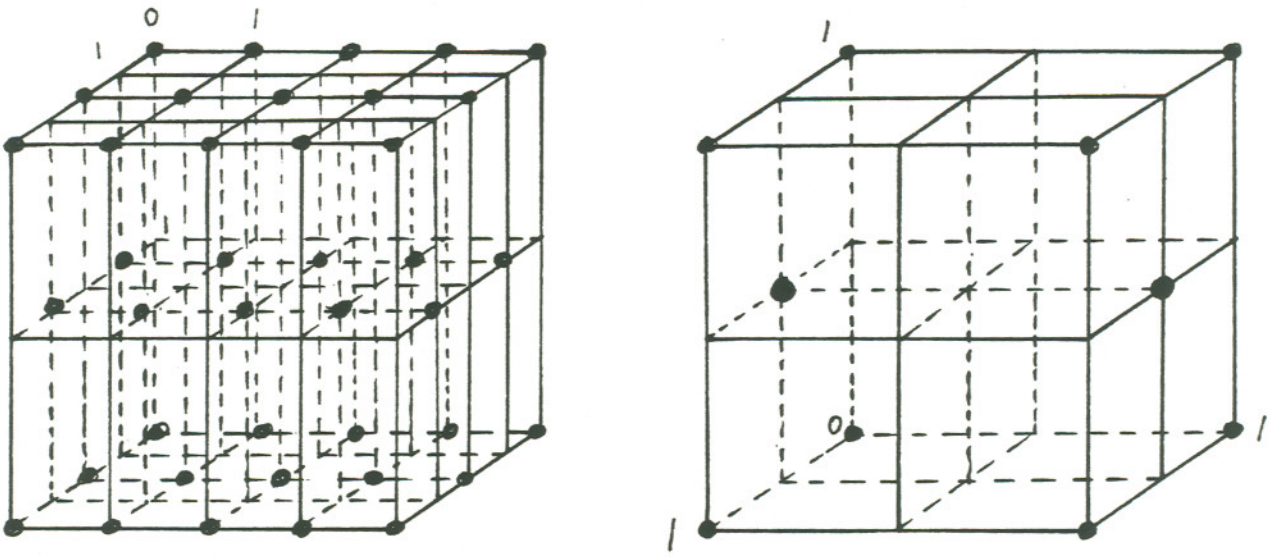


reciprocal lattice





(a) Progressive scanning



$$T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

$\gamma_{[0\,0\,0]^t} = 1$

$\gamma_{[0\,1\,1]^t} = 1$

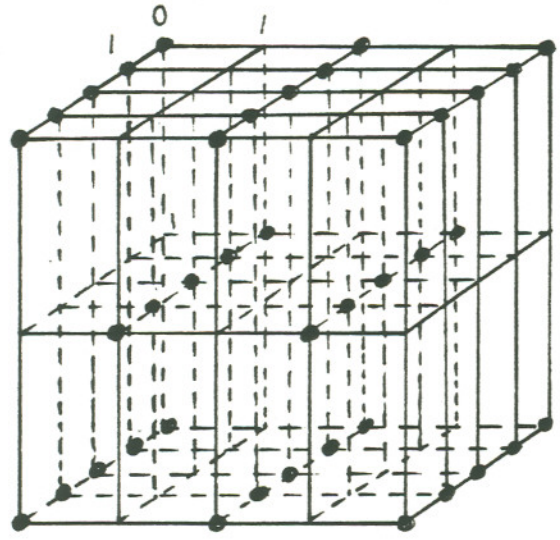
$\Gamma_{[0\,0\,0]^t} = 2$

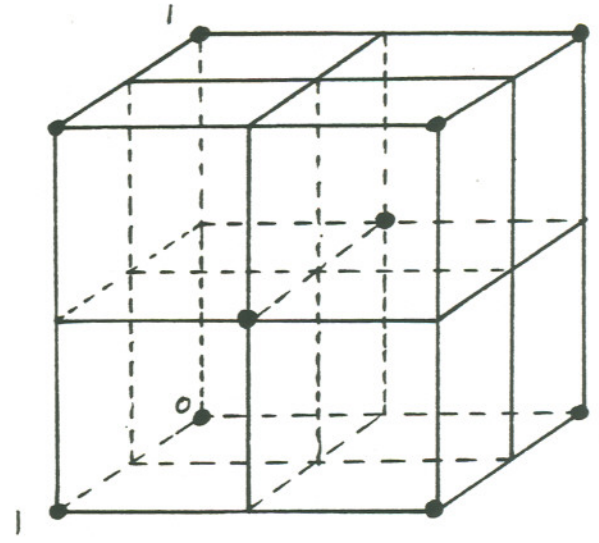
$\Gamma_{[0\,1\,1]^t} = 2$

Additional repeat position:

(0,0,0) and $(0,\frac{1}{2},\frac{1}{2})$

(b) Interlaced scanning





$$T = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$\gamma_{[0\,0\,0]^t} = 1$

$\gamma_{[1\,1\,0]^t} = 1$

$\Gamma_{[0\,0\,0]^t} = 2$

$\Gamma_{[0\,0\,1]^t} = 2$

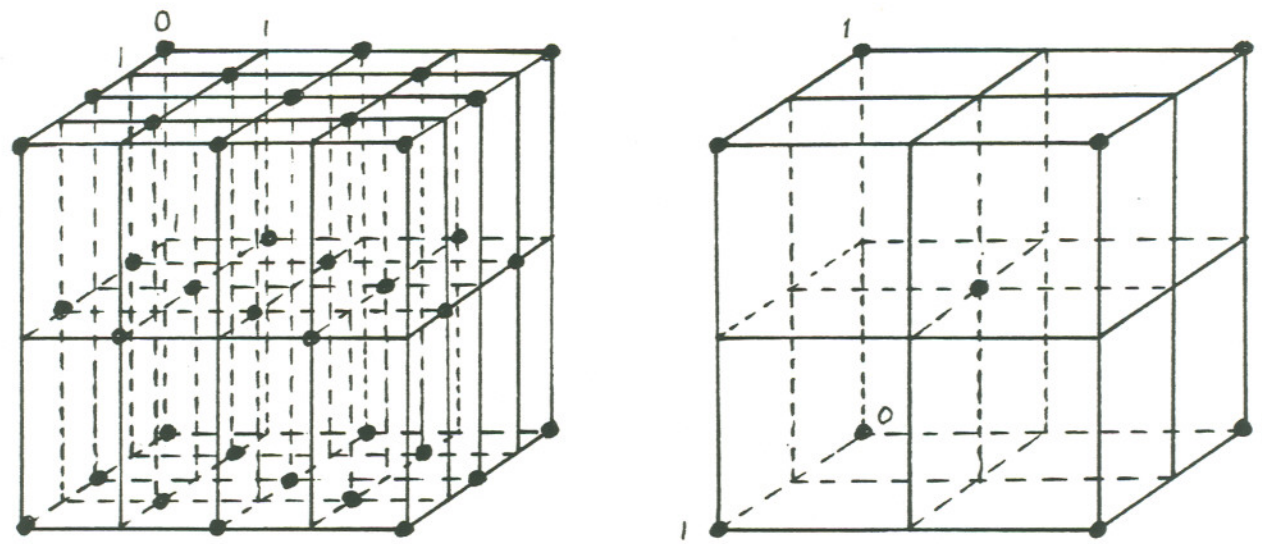
$\Gamma_{[1\,1\,0]^t} = 2$

$\Gamma_{[1\,1\,1]^t} = 2$

Additional repeat position:

(0,0,0) and $(\frac{1}{2},\frac{1}{2},0)$

(c) Offset sampling



$$T = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

$\gamma_{[\,0\,0\,0\,]^t} = 1$

$\gamma_{[\,1\,0\,1\,]^t} = 1$

$\gamma_{[\,1\,1\,0\,]^t} = 1$

$\gamma_{[\,0\,1\,1\,]^t} = 1$

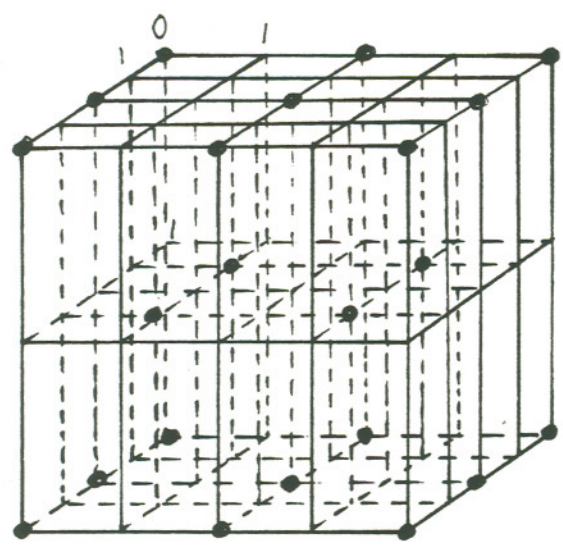
$\Gamma_{[\,0\,0\,0\,]^t} = 4$

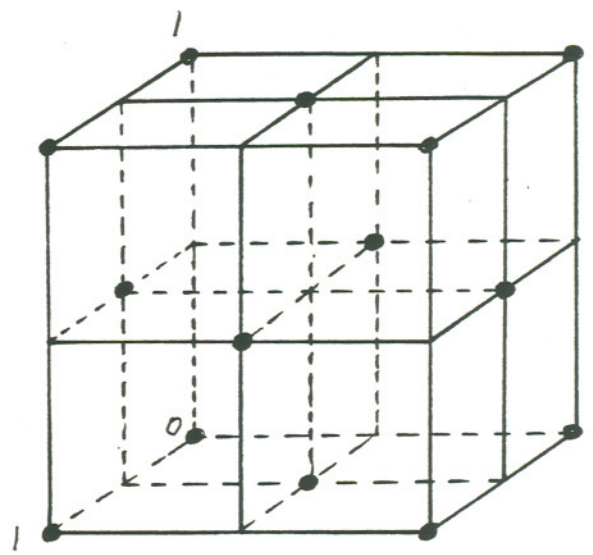
$\Gamma_{[\,1\,1\,1\,]^t} = 4$

Additional repeat position:

(0,0,0) and $(\frac{1}{2},\frac{1}{2},\frac{1}{2})$

(d) Offset sampling with field offset





$$T = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

$\gamma_{[0\,0\,0]^t} = 1$

$\gamma_{[1\,1\,1]^t} = 1$

$\Gamma_{[0\,0\,0]^t} = 2$

$\Gamma_{[0\,1\,1]^t} = 2$

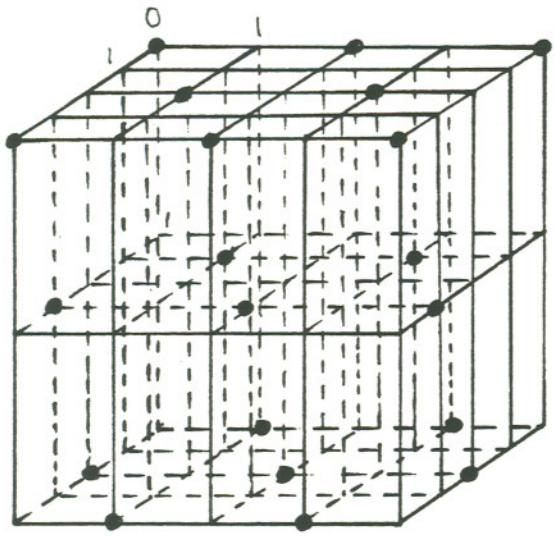
$\Gamma_{[1\,0\,1]^t} = 2$

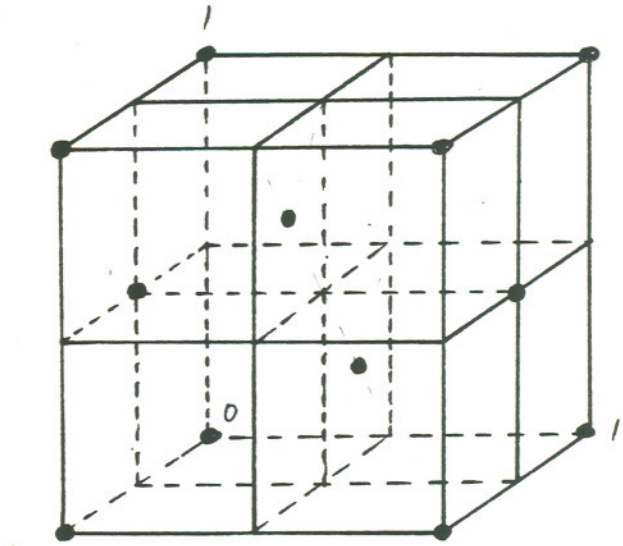
$\Gamma_{[1\,1\,0]^t} = 2$

Additional repeat position:

(0,0,0), $(0,\frac{1}{2},\frac{1}{2})$, $(\frac{1}{2},0,\frac{1}{2})$ and $(\frac{1}{2},\frac{1}{2},0)$

(e) Field quincunx





$$T = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{bmatrix}$$

$\gamma_{[000]^t} = 1$ $\qquad$ $\Gamma_{[000]^t} = 8$

$\gamma_{[111]^t} = 1$ $\qquad$ $\Gamma_{[022]^t} = 8$

$\gamma_{[102]^t} = 1$ $\qquad$ $\Gamma_{[111]^t} = 8$

$\gamma_{[013]^t} = 1$ $\qquad$ $\Gamma_{[133]^t} = 8$

$\gamma_{[120]^t} = 1$ $\qquad$ $\Gamma_{[200]^t} = 8$

$\gamma_{[031]^t} = 1$ $\qquad$ $\Gamma_{[222]^t} = 8$

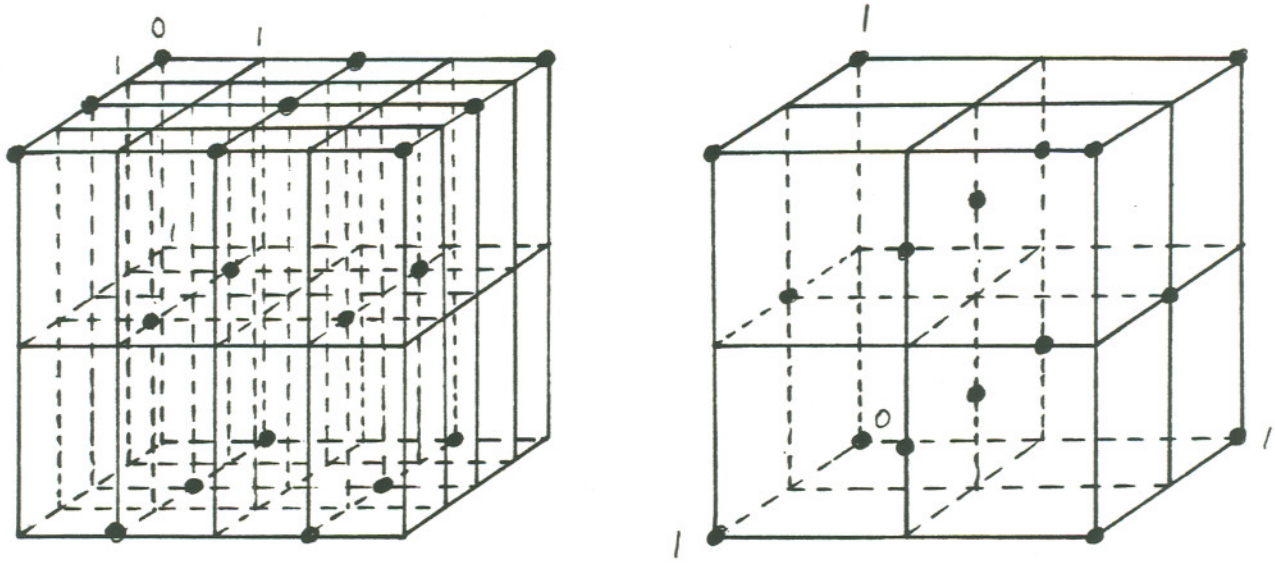
$\gamma_{[022]^t} = 1$ $\qquad$ $\Gamma_{[311]^t} = 8$

$\gamma_{[133]^t} = 1$ $\qquad$ $\Gamma_{[333]^t} = 8$

Additional repeat position:

(0,0,0), $(0,\frac{1}{2},\frac{1}{2})$, $(\frac{1}{2},\frac{1}{4},\frac{1}{4})$ and $(\frac{1}{2},\frac{3}{4},\frac{3}{4})$

(f) Line quincunx



$$T = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

$\gamma_{[000]^t} = 1$

$\gamma_{[120]^t} = 1$

$\gamma_{[111]^t} = 1$

$\gamma_{[031]^t} = 1$

$\Gamma_{[000]^t} = 4$

$\Gamma_{[002]^t} = 4$

$\Gamma_{[021]^t} = 4$

$\Gamma_{[023]^t} = 4$

$\Gamma_{[110]^t} = 2+2j$

$\Gamma_{[111]^t} = 2-2j$

$\Gamma_{[112]^t} = 2+2j$

$\Gamma_{[113]^t} = 2-2j$

$\Gamma_{[130]^t} = 2-2j$

$\Gamma_{[131]^t} = 2+2j$

$\Gamma_{[132]^t} = 2-2j$

$\Gamma_{[133]^t} = 2+2j$

$\Gamma_{[200]^t} = 4$

$\Gamma_{[202]^t} = 4$

$\Gamma_{[221]^t} = 4$

$\Gamma_{[223]^t} = 4$

$\Gamma_{[310]^t} = 2+2j$

$\Gamma_{[311]^t} = 2-2j$

$\Gamma_{[312]^t} = 2+2j$

$\Gamma_{[313]^t} = 2-2j$

$\Gamma_{[330]^t} = 2-2j$

$\Gamma_{[331]^t} = 2+2j$

$\Gamma_{[332]^t} = 2-2j$

$\Gamma_{[333]^t} = 2+2j$

Additional repeat position:

$(0,0,0)$, $(0,\frac{1}{2},\frac{1}{2})$, $(\frac{1}{2},\frac{1}{4},0)$, $(\frac{1}{2},\frac{1}{4},\frac{1}{2})$, $(\frac{1}{2},\frac{3}{4},0)$ and $(\frac{1}{2},\frac{3}{4},\frac{1}{2})$

(g) Modified line quincunx

Fig.2.15 Signal and spectrum of typical samplings

In summary, we can find the repeat positions by following the procedure below:

1. Choose periodic block (γ_i) in lattice, translation matrix $\underline{T}$ and scale matrix $\underline{L}$
2. Determine Γ_i by utilizing Eq.(2.26)
3. Calculate $(\underline{TL})^{-t}i$ with respect to non zero Γ_i and select the ones in $\underline{I}$

The importance of the repeat position is to confine the spectrum of the signal to avoid aliasing. As above, several commonly used sampling pattern with their repeat positions are given in Fig.2.15.

2.3. TV Spectra

2.3.1. Three Dimensional Spectrum of the NTSC System

In the early 1950's, people were not satisfied with the black and white (or monochrome) television, so various systems were adopted to introduce color television. NTSC system[8] is one of them. Since NTSC is successful both commercially and technically, it plays an important role not only in today's television, but also in the future's compatible HDTV (High Definition TeleVision). Many countries use this system, U.S.A, Japan etc. So, before we start discussing the spectrum of NTSC signal, let us introduce the NTSC system at first.

2.3.1.1. NTSC System

For black and white television system, the agreed upon standards were 525 lines per frame, 59.94 or 60 (it will be explained later) fields per second, 2:1 interlaced scan and 4:3 aspect ratio. Channel spacing for broadcasting was set at 6.0 MHz. Video signal was modulated by VSB-AM (Vestigial SideBand Amplitude Modulation). The sound carrier frequency was set at 4.5 MHz. The color system was introduced based on human vision which indicates that excellent color can be reproduced by superimposing three essential pure colors red, green and blue. Moreover, considering compatibility with black and white systems and reducing noise deterioration, the luminance signal (Y) was maintained and color information was added into the black and white signal by inserting a subcarrier modulated in quadrature by two color difference signals I and Q whose bandwidth are about 1.5 MHz and 0.5 MHz respectively. Y, I and Q were chosen as the following form

$$Y = 0.59G + 0.11B + 0.30R \tag{2.30a}$$

$$I = -0.27(B-Y) + 0.74(R-Y) \tag{2.30b}$$

$$Q = 0.41(B-Y) + 0.48(R-Y) \tag{2.30c}$$

In order to minimize the appearance of the subcarrier in the picture, the color subcarrier frequency (f_{sc}) was chosen to be an odd multiple of one half the horizontal scanning frequency (f_h), which is given by

$$f_{sc} = \frac{(2n+1)}{2} f_h \tag{2.31}$$

For 525 lines, 60 Hz system, f_h is 15,750 Hz, then

$$f_{sc} = \frac{455}{2} f_h \approx 3.58 \text{ MHz}$$

To avoid the mutual interference between the color subcarrier and the sound carrier (f_s), the color subcarrier was interleaved with the sound carrier. Since the sound carrier was used, the horizontal scanning frequency was changed so that the sound carrier frequency would be an even multiple of the horizontal scanning frequency.

$$f_s/f_h = 4{,}500{,}000/15{,}750 = 285.71 \approx 286$$

$$f_h = \frac{f_s}{286} = 15{,}734.26 \text{ Hz}$$

$$f_f(\text{field}) \equiv f_v(\text{vertical}) = \frac{2}{525} f_h = 59.94 \text{ Hz}$$

$$f_{sc} = \frac{455}{2} f_h \approx 3.58 \text{ MHz}$$

That is the reason the NTSC system in U.S.A. uses 59.94 Hz as its field frequency. Fig.2.16 gives the NTSC signal distribution in the frequency domain and Fig.2.17 describes the color subcarrier procedures with very quick flybacks.

51

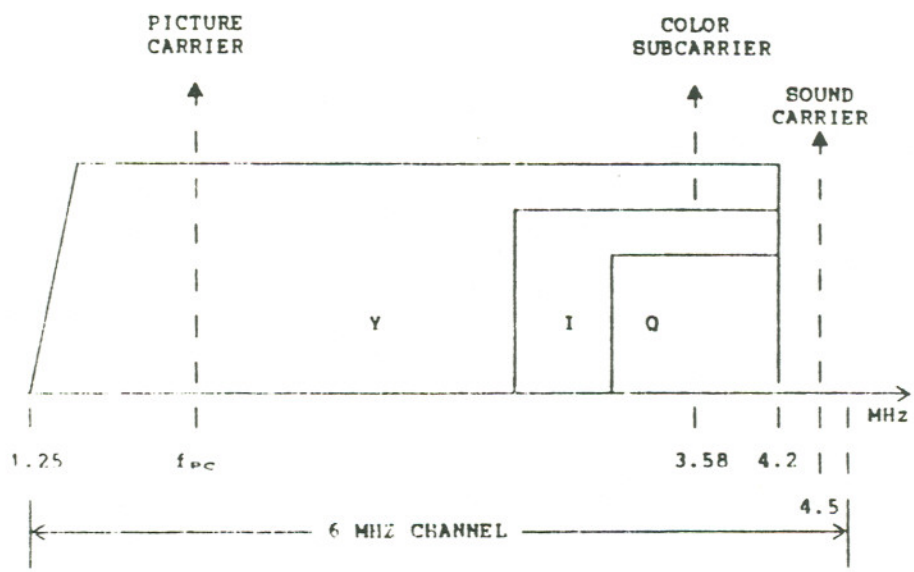
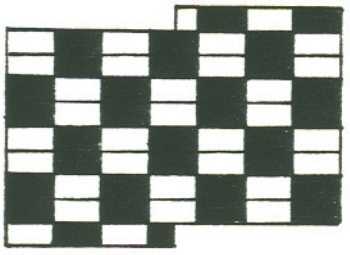
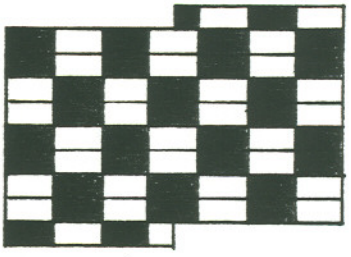


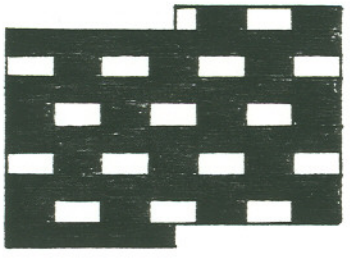
Fig.2.16 NTSC signal distribution in frequency domain



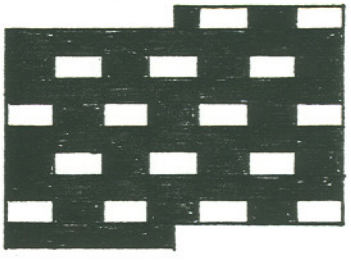
frame (picture) 1



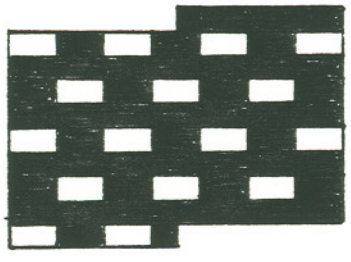
frame (picture) 2



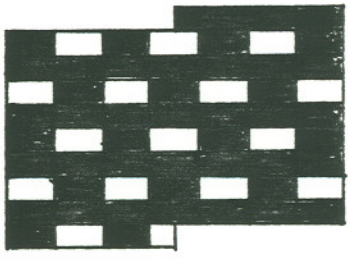
field 1 of picture 1



field 1 of picture 2



field 2 of picture 1



field 2 of picture 2

(a) (b)

Fig.2.17 Colour subcarrier procedures by example of dot interlacing

2.3.1.2. 3-D Spectrum of NTSC System

The three dimensional spectrum of NTSC system[9] is illustrated in Fig.2.18 by the assumption of system specification as 525 lines, 4.2 MHz bandwidth and 59.94 Hz field frequency.

Considering the video information confined in a rectangular box A, B and C in x, y and t space, the first harmonics are corresponding to $\frac{1}{A}, \frac{1}{B}$ and $\frac{1}{C}$ respectively. In the case of pixel description, the video signal becomes periodic. Accordingly, a horizontal spatial frequency f_x is a harmonic of one cycle per picture width corresponding to $\frac{n}{A}$ where n is an integer; a vertical spatial frequency f_y is a harmonic of one cycle per picture height corresponding to $\frac{m}{B}$ where m is an integer; and f_t is a harmonic of one cycle per chosen time interval.

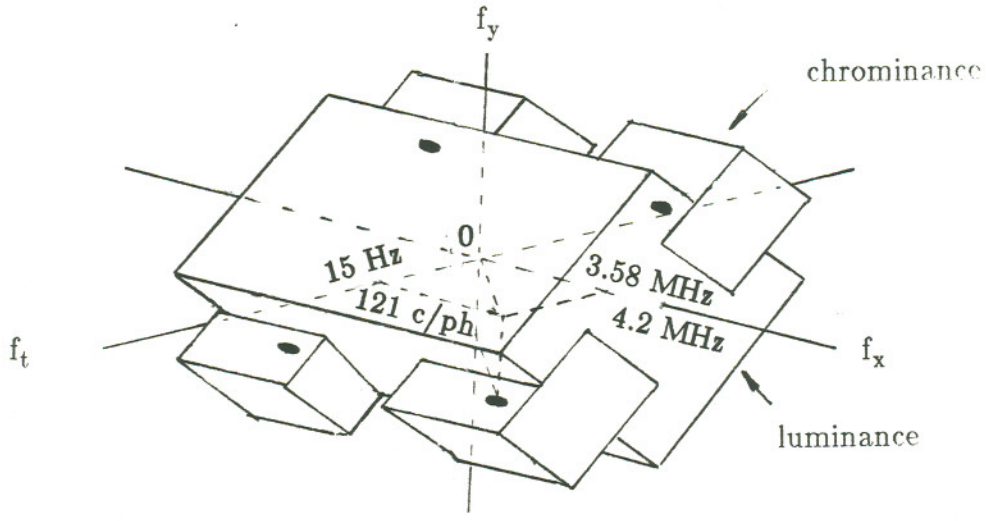
Using mathematical tools, the spatio-temporal frequencies are conveniently described in the complex exponential form, i.e. $e^{j2\pi(f_x x + f_y y + f_t t)}$. Recall the scanning scheme, the path of the scanning spot in x, y and t space is given by

$$x = A\, f_h\, t$$

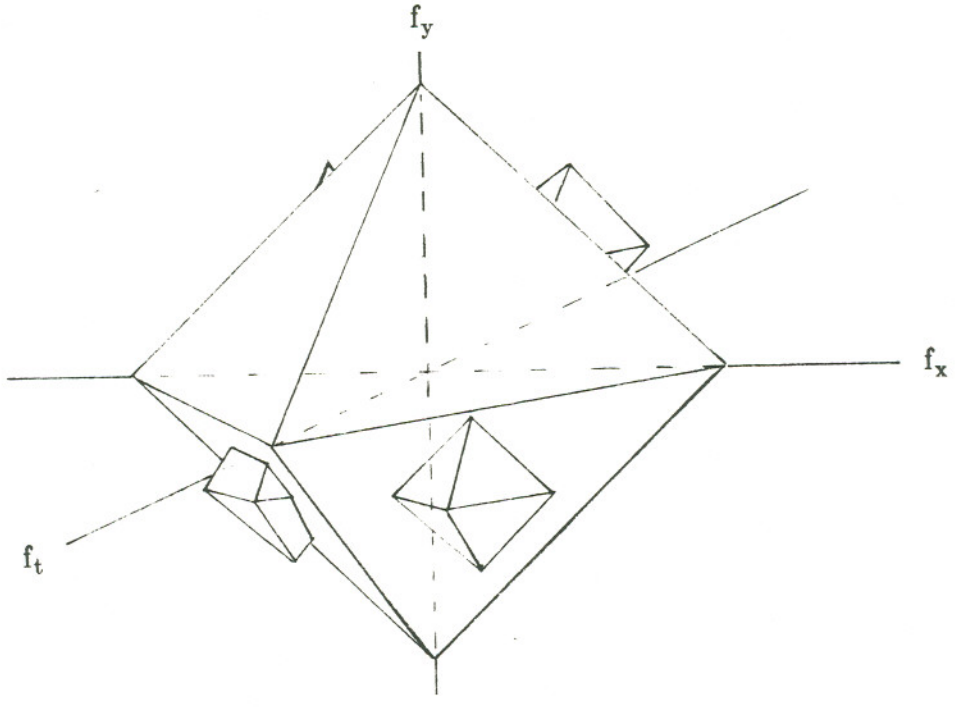
$$y = -B\, f_v\, t \qquad \text{(a negative sign is due to top to bottom scanning)}$$

$$t = C\, f_t\, t \tag{2.32}$$

here A and B correspond to unit width or unit height.



(a) Three dimensional spectrum of NTSC signal



(b) Three dimensional quincunx sampling spectrum of NTSC signal

Fig.2.18 Three dimensional spectrum of NTSC signal

Since f_x, f_y and f_t are harmonics individually, we get

$$e^{j2\pi ft} \equiv e^{j2\pi(f_x x + f_y y + f_t t)}$$

$$= e^{j2\pi(\frac{n}{A}Af_h t - \frac{m}{B}Bf_v t + f_t t)}$$

$$= e^{j2\pi(nf_h - mf_v + f_t)t}$$

$$f = nf_h - mf_v + f_t \tag{2.33}$$

Therefore, a spatio-temporal frequency $(f_x,f_y,f_t) = (\frac{n}{A},\frac{m}{B},f_t)$ corresponds to the video frequency $nf_h - mf_v + f_t$. Any spatio-temporal frequency of the form $(\frac{n}{A},\frac{m}{B},f_t) = (\frac{n}{A},\frac{m}{B},mf_v - nf_h)$ is mapped to zero video frequency which causes aliasing. To avoid aliasing, the spatio-temporal components have to be confined near the origin. The 525 vertical samples can support frequencies of up to 262.5 c/ph. The horizontal frequency limits correspond to the 4.2 MHz bandwidth or 266.93 c/pw. The temporal boundary of the region is based on frame frequency i.e. 29.97 Hz. The diamond shape of the band in the vertical temporal frequency plane is due to the line interlaced scanning used in NTSC.

Recall $f = nf_h - mf_v + f_t$, then the color subcarrier frequency is given by

$$f_{sc} = \frac{455}{2}f_h = 227\ f_h + 131\ f_v + \frac{f_v}{4}$$

or

$$f_{sc} = \frac{455}{2} f_h = 228\, f_h - 131\, f_v - \frac{f_v}{4}$$

Therefore, it is located at $(\frac{227}{A}, \frac{-131}{B}, \frac{f_v}{4})$ and $(\frac{228}{A}, \frac{131}{B}, -\frac{f_v}{4})$ in 3-D frequency domain. Since the video signal is considered as real, its Fourier transform is symmetric with respect to the origin.

$$F(f_x,f_y,f_t) = \int\int\int f(x,y,t)\, e^{-j2\pi(xf_x+yf_y+tf_t)}\, dxdydt = F^*(-f_x,-f_y,-f_t)$$

So, there are the components at $(-\frac{227}{A}, \frac{131}{B}, -\frac{f_v}{4})$ and $(-\frac{228}{A}, -\frac{131}{B}, \frac{f_v}{4})$ which corresponds to $-f_{sc}$. Assume $A = 1$ and $B = 1$, the color subcarrier can be located only at the four of the eight corners of 3-D frequency space given by

(227 c/pw , −131 c/ph , 14.985 Hz)

(−227 c/pw , 131 c/ph , −14.985 Hz)

(228 c/pw , 131 c/ph , −14.985 Hz)

(−228 c/pw , −131 c/ph , 14.985 Hz)

In practice, Kell factor etc. results NTSC system to have an active picture height as 484 lines out of 525 line system and a horizontal active line period equal to 0.827 of total line period.

$\frac{1}{B}$ is approximated by $\frac{484}{525} \approx 0.922$;

$\frac{1}{A} \approx 0.827$ (corresponding to c/pw)

$$= 0.827 \times \frac{4}{3} \approx 0.620 \text{ (corresponding to c/ph);}$$

and f_t is maintained to be 59.94 Hz.

The color subcarrier is located at these positions and shown in Fig.2.18.

(140.74 c/ph , −120.78 c/ph , 14.985 Hz)

(−140.74 c/ph , 120.78 c/ph , −14.985 Hz)

(141.36 c/ph , 120.78 c/ph , −14.985 Hz)

(−141.36 c/ph , −120.78 c/ph , 14.985 Hz)

The chrominance spectrum then is centered around four subcarrier components. The psychophysical evidence regarding the color perception indicates that the human visual system is less sensitive to high chrominance frequencies than to high luminance frequencies. This leads to the compression of the color signal bandwidth to more than half of the luminance signal. The chrominance spectrum can be confined in 121 c/ph in vertical direction, 1.5 MHz (59.11 c/ph) and 0.5 MHz (19.70 c/ph) for I and Q respectively in horizontal direction and $\frac{f_f}{8} = \frac{59.94}{8} \approx 7.493$ Hz in temporal direction (Fig.2.18(a)). The optional spectrum can be also selected as pyramid shape if sampling such as quincunx is adopted both in vertical and horizontal direction (Fig.2.18(b)).

2.3.2. Three Dimensional Spectrum of the PAL System

As mentioned in NTSC system, color information was introduced by the color subcarrier. Wide luminance bandwidth results the optimum color subcarrier frequency to be at 3.6 MHz. However, with a 4 MHz video channel, the modulating frequencies would be necessarily single sideband and crosstalk between R-Y and B-Y would result. To overcome this crosstalk problem, two methods were suggested. One is IQ working in today's NTSC system, the other is CPA (Color Phase Alternation). PAL system[8] is one of such phase tolerant system.

2.3.2.1. PAL System

Compare with NTSC, the PAL chroma signal is nothing but alternating the sign of R-Y occurring line by line of each field, which is given by

$$C_{PAL} = \frac{B-Y}{2.03}\sin\omega_s t \pm \frac{R-Y}{1.14}\cos\omega_s t$$

$$= U\sin\omega_s t \pm V\cos\omega_s t \qquad (2.32)$$

where ω_s is a subcarrier angular frequency. This forms two subcarriers U and V. The switching is carred out on V rather than U because the latter has the larger weighting factor (2.03>1.14) and peak-to-peak swing. The bandwidths of U and V can be now equal and the same as that of the NTSC I bandwidth (1.5 MHz). The subcarrier frequency f is selected to be quarter-line plus $f_v/2$ offset which can

not use the simple NTSC half-line offset because of cross-talk between the V and luminance components. In this system there are 625 lines, 5.5 MHz bandwidth, 50 Hz field rate and 2:1 interlaced PAL system. As the result, the U subcarrier is $283\frac{3}{4}f_h + \frac{1}{2}f_v$. And the V carrier can be obtained by shifting U by $f_h/2$.

2.3.2.2. 3-D Spectrum of PAL System

In this system, the 625 lines can support the frequencies up to 312.5 c/ph by Nyquist limit; 5.5 MHz bandwidth can give a maximum horizontal resolution of $\frac{f}{f_h} = 352$ c/pw ($f_h = 15{,}625$ Hz) and 50 Hz field rate leads to a 25 Hz maximum temporal resolution. Similar to NTSC, the diamond shape in the vertical-temporal frequency plane is used to decrease data rate. The two subcarriers can be written as

$$f_U = 283\frac{3}{4} f_h + \frac{f_v}{2}$$

$$= 284\, f_h - 78\, f_v + \frac{3}{8} f_v$$

$$= 283\, f_h + 235\, f_v - \frac{1}{8} f_v$$

$$f_V = f_U - \frac{f_h}{2}$$

$$= 284\, f_h - 234\, f_v + \frac{1}{8} f_v$$

$$= 283\ f_h + 78\ f_v - \frac{3}{8}\ f_v$$

so that U is located at

(284 c/pw , 78 c/ph , 18.75 Hz)

(−284 c/pw , −78 c/ph , −18.75 Hz)

(283 c/pw , −235 c/ph , −6.25 Hz)

(−283 c/pw , 235 c/ph , 6.25 Hz)

and V is positioned at

(284 c/pw , 234 c/ph , 6.25 Hz)

(−284 c/pw , −234 c/ph , −6.25 Hz)

(283 c/pw , −78 c/ph , −18.75 Hz)

(−283 c/pw , 78 c/ph , 18.75 Hz)

The color information is confined in the box of 1.5 MHz (96 c/pw) in horizontal, 156.25 c/ph (625/4) in vertical and 6.25 Hz (50/8) in temporal. The 3-D spectrum of PAL system is demonstrated in Fig.2.19[10].

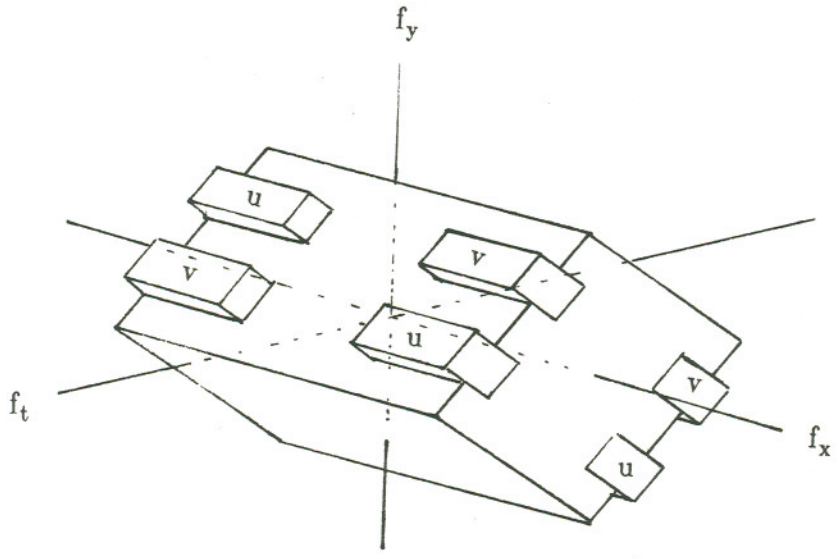


Fig.2.19 Three dimensional spectrum of PAL signal

HDTV PROPOSALS

In the early 1950's, the black and white television was not satisfactory, and the color television was invented. Now, a desire to further improve the picture quality is resulting into the advanced television systems which will meet the demands of getting a more realistic viewing experience than is offered by today's TV. So, HDTV was introduced, and it has become a hot topic recently. The aim of HDTV is to enlarge the aspect ratio (widescreen presentation on a display), to increase the resolution both in horizontal and vertical direction, to improve the picture quality (such as flickering) and to get higher audio quality. However, it is not so easy to make these changes. The first focus is on the compatibility[11] with today's systems. The price of HDTV system must be reasonable so as to make customers comfortable. A need exists to set up a single world-wide production standard instead of various systems existing now such as NTSC, PAL and SECAM (sequential color signals and a memory device --- sequential à mémoire). Honestly, there might be an opportunity to have a revolutionary change from today's multiple standards. A major problem is how to make a compromised choice to improve quality of picture, for example, when spatial resolution increases, sometimes it will result in a loss of temporal resolution. To meet the needs of higher quality, some fairly complex signal processing techniques such as the bandwidth compression and the subsequent processing at the display will be used. Until recently, this has been far too difficult and expen-

sive[2]. However, the rapid development of digital technology and VLSI has now made it possible. So far, a number of proposals[11–50] have been presented. Since U.S.A. TV network is established on NTSC system, here we only mention some proposals pertaining to NTSC system.

3.1. Defects of NTSC

The commercial and technological success of the NTSC system is undeniable. With the development of technology, both professionals and lay men have become more critical about the defects of today's NTSC system. The main flaws of this system are described below.

Limited vertical resolution ----- The Kell factor[51,52] and active interlaced scanning factor lead to unsatisfactory vertical resolution out of scanning system.

Limited horizontal resolution ----- 4.2 MHz bandwidth decides the horizontal resolution to be approximately equal to vertical resolution.

Limited chrominance resolution ----- With IQ working scheme, Q is limited to a 500 kHz bandwidth and I is up to 1.5 MHz bandwidth. It is by no means an excellent color resolution since they are not RGB primaries and their variations might introduce image degradation.

Small aspect ratio ----- It is demonstrated that the aspect ratio which has a better match to the human visual system, is 5:3, 16:9 or even 2:1 instead of today's 4:3 ratio.

Cross color ----- This is the effect of misinterpreting luminance signal into false color information.

Cross luminance ----- This is the result of artifacts of luminance produced by the color signal.

Temporal aliasing ----- This is the effect of moving picture going too fast for displayer to follow. For example, if a pendulum in the picture oscillates at 60 fields/sec sampling rate, it appears stationary.

Large area flicker ----- Due to the short-persistence phosphor used in the modern cathode-ray tube (CRT) displays, a picture vanishes very rapidly in brightness if the refresh frequency is low. That generates large area flicker. Studies show that although the 59.94 Hz NTSC field rate is satisfactory for most, 100 Hz is clearly high enough to perceive the picture without flashing for even the most critical.

Interline flicker ----- This is the effect that flashing on the screen will happen at the horizontal borders of objects with interlaced scan.

3.2. Developed Methods on Improving Quality of Pictures

3.2.1. Diagonal Filtering and Subsampling

There is psychophysical evidence regarding to "oblique effect" which indicates that human visual system is less sensitive to high diagonal spatial frequencies than high vertical or horizontal frequencies. This leads to the idea of

diagonal filtering and subsampling. As discussed in Section 2.2.4.2, various sampling patterns are given, of which quincunx pattern is more useful.

3.2.2. Pre- and Post- Combing

Since the artifacts will be caused by mutual interference between high frequency luminance information and the color signal due to the overlapped spectra, filtering the signal before modulation will eliminate aliasing spectra. Because the color subcarrier frequency is an odd multiple of one half the horizontal scanning frequency, we can use a comb filter to separate the luminance and chrominance signals[12]. Due to the fact that the color subcarrier (Fig.2.17) has a 180° phase shift line by line, field by field and frame by frame, if the signal is in phase from one line to the next line or from one field to the next field or from one frame to the next frame, it is assumed to be luminance information; if the signal is out of phase from one line to the next line or from one field to the next field or from one frame to the next frame, it is assumed to be color information. Such filters are called line comb filter, field comb filter and frame comb filter, respectively. The usage of these filters depends on the motion and resolution of the picture. Generally speaking, frame comb filter is used for stationary picture, while line comb filter is used for moving picture. It will be discussed in detail later.

3.2.3. Fukinuki Hole and Proposal

As mentioned in the above, color information interleaved with the luminance information. Fukinuki found out that a portion of the spectrum devoted

to color is poorly used and there is no color information in four quadrants, which is named Fukinuki hole[13,14] as shown in Fig.2.18(a). He gave the idea that high resolution luminance information can be inserted into these four quadrants with the another phase shift given in Fig.3.1. The spectrum of this proposal looks like the one shown in Fig.3.2. This technique produces motion defects. And motion compensation circuitry must be included in the receiver.

3.2.4. Quadrature Modulation of the Picture Carrier

The term QUME[15] standing for QUadrature Modulating Extended is used to increase the horizontal resolution or to increase the aspect ratio. The QUME is produced by VSB modulation of the two modulated signals using carriers with 90° phase difference. One signal is the normal NTSC signal, and the other is the additional signal which is processed by the band limited Nyquist filter in order to decrease crosstalk. Fig.3.3(d) gives the one dimensional spectrum of the transmission signal.

3.2.5. Motion Compensation

To deal with the problem of high spatial detail in a moving area, interpolation with motion compensation[5] will be used. The simplest one is based on a rigid moving body assumption which presumes the rigid body moves uniformly between input fields with only translation motion. Therefore, if the motion vector of the block is given, the block position can be easily determined by simple linear interpolation method as shown in Fig.3.4.

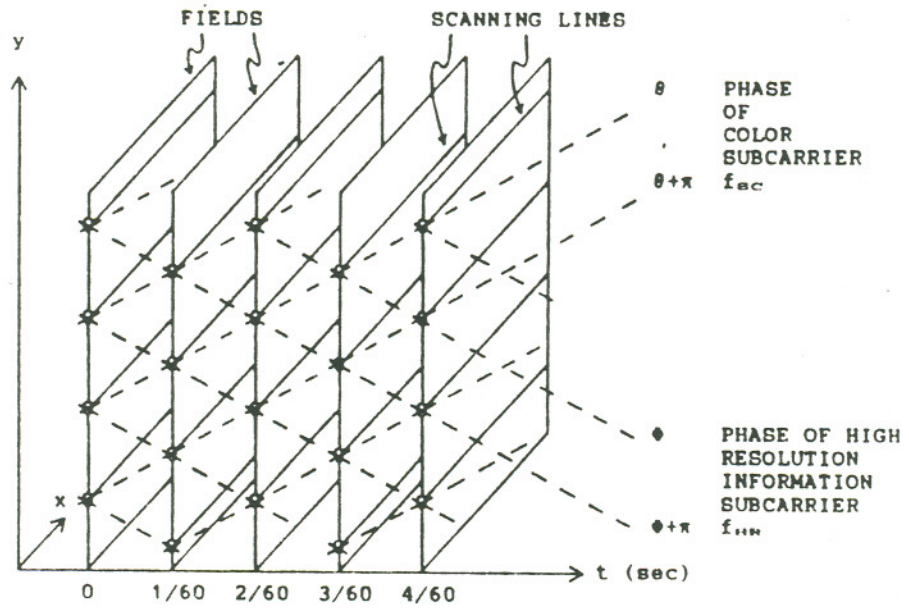


Fig.3.1 Fukinuki proposal

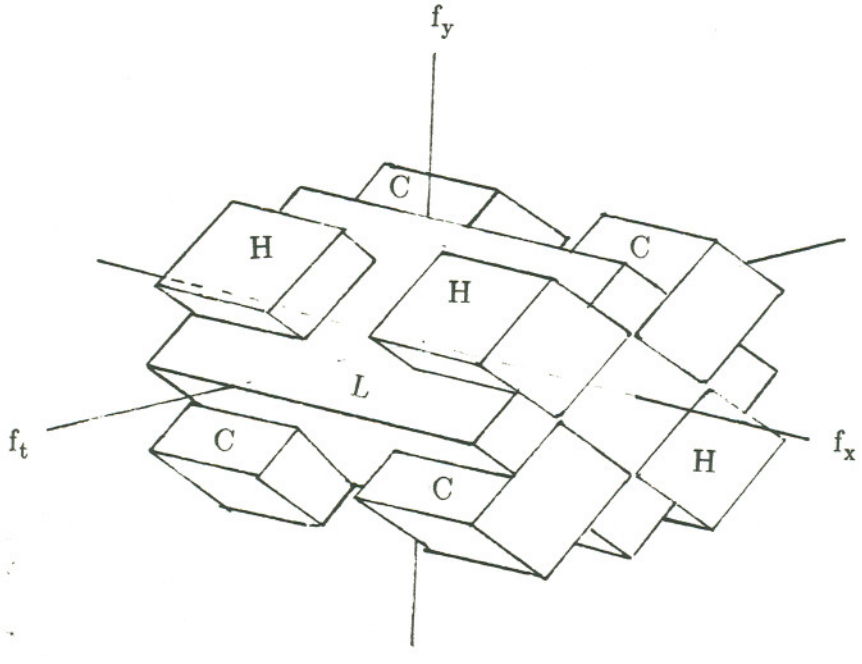


Fig.3.2 Three dimensional spectrum of Fukinuki porposal

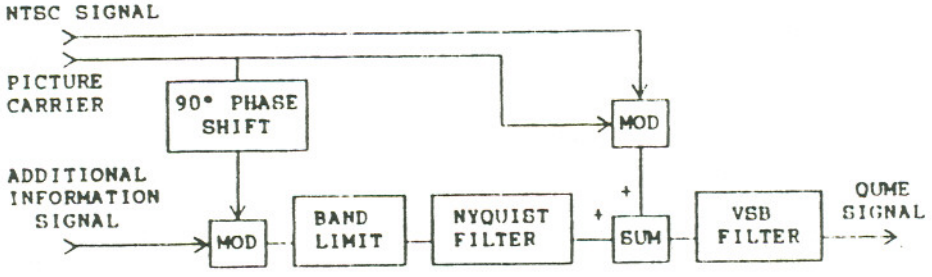


FIGURE A QUME MODULATING CIRCUIT

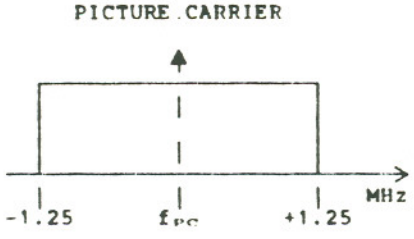


FIGURE B BAND LIMIT FILTER CHARACTERISTIC

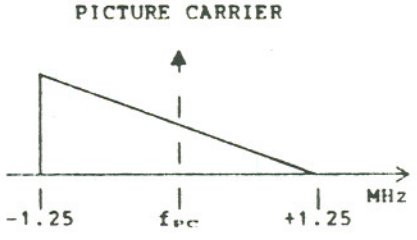


FIGURE C NYQUIST FILTER CHARACTERISTIC

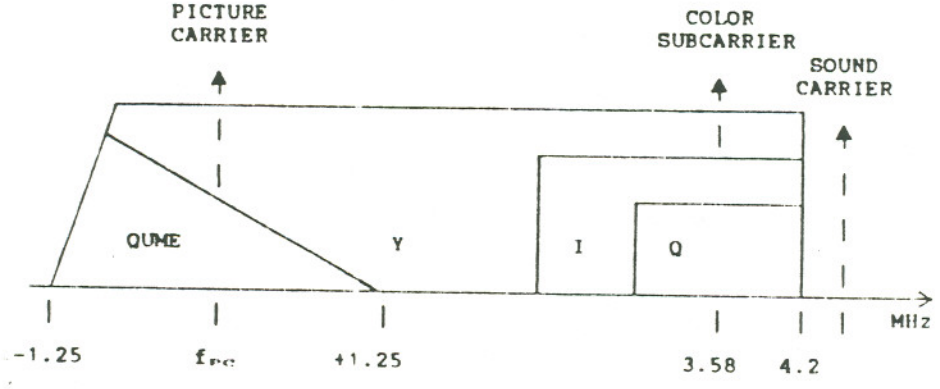


FIGURE D SPECTRUM OF THE QUME SIGNAL

Fig.3.3 QUME process

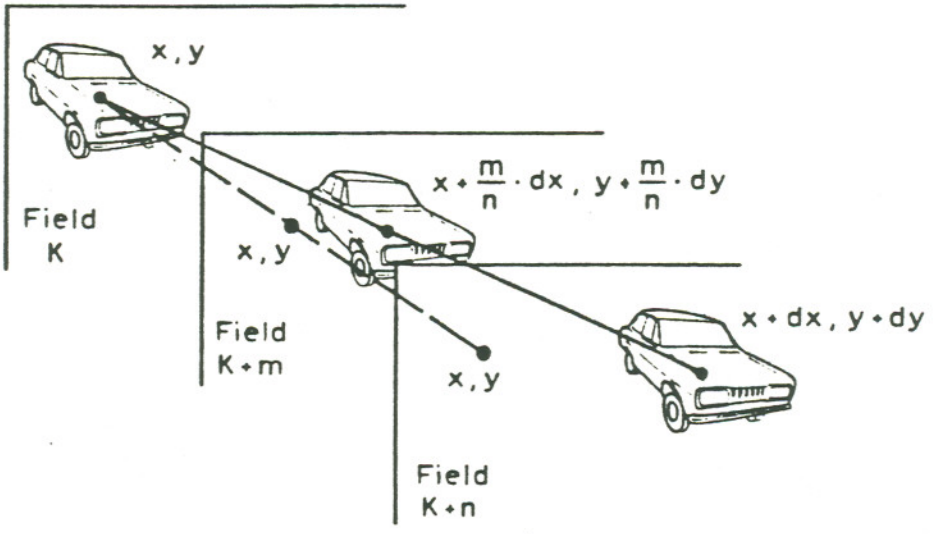


Fig.3.4 Motion compensted field interpolation

3.2.6. Deinterlacing

Obviously, a progressive scan system can produce a better picture than an interlaced scan system, for instance, field aliasing can be eliminated. It doubles the bandwidth and introduces a penalty in signal to noise ratio than that of interlaced scan system. However, if we use an interlaced scan system, the vertical resolution is decreased by a factor of two as that of the progressive scan system. How can we obtain a better picture without the wider bandwidth? The way to implement this is to use an interpolation technique in the interlaced scan system. The easiest way is to store one line, double the horizontal scanning frequency and display every scan line twice during the normal scan line period. Evidently, diagonal lines in the picture with this display method become distorted. Another way is to require two line storage and insert the average of two adjacent scan lines between these two lines. The distortion of diagonal lines is decreased. If a field store is used to display the "missing" line, it is very good for still picture but motion compensation circuitry is required to take care of the motion in the picture.

3.2.7. Flicker Elimination

High refresh rate display can reduce the effect of flashing. Three schemes (Fig.3.5)[5,16] are often used for this purpose. Temporal averaging can smooth the picture but edge blur occurs. Field repetition can be used for moving pictures. It removes large area flicker but not interline flicker. Picture repetition removes both large area flicker and interline flicker. But it can only be used in the still

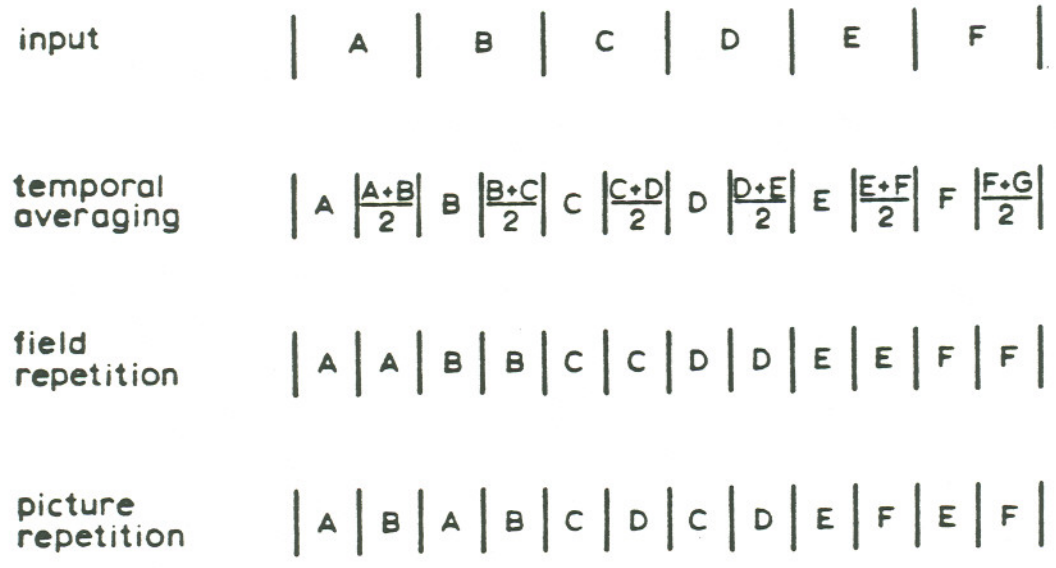


Fig.3.5 High rate scanning

picture since in the moving zone it will give rise to clearly visible degradation.

3.3. ATV (Advanced TV) Proposals

From a survey of the various rosy proposals for HDTV[17], you will find that the following three format systems[1] are included:

* New satellite and terrestrial system. The new transmission system are not compatible with today existing systems. It requires a transcoding coverter so as to serve both HDTV receivers and conventional receivers. HDTV receivers can display the full quality HDTV picture while conventional receivers might display HDTV signals but with less quality.

* Compatible system. The compatible transmission systems generally transmit two signals: one is a compatible signal with conventional quality for today's existing systems and the other is an augmentation signal with additional information for HDTV systems. HDTV receivers can receive both conventional format and full quality HDTV signals while conventional receivers can display only the compatible signals.

* Multistep systems[18]. In the first step, the system provide an improved quality signal. It allows conventional receivers to display the picture with conventional quality while HDTV receivers (HDTV-1) display with full quality. In the second step, a system fully compatible with the earlier format in the first step takes over. Then conventional transmission will be phased out. HDTV-2

receivers will display the full quality picture while HDTV-1 receivers operate with degraded quality.

Here, we only introduce a few proposals in brief.

In Europe a new system called MAC system (Multiplexed Analogue Components) is proposed. In this system, the luminance and color difference and multiple digital sound signals are compressed in time and placed on the same signal line. The system can offer high performance than composite system, with full bandwidth luminance and lack of crosstalk because of separation of luminance and color difference signals in sequence. However, color vertical resolution is lower because of the line sequential format. Also, baseband bandwidth is increased by the luminance compression ratio. Different compression ratios, data rates and the number of sound channels result in different MAC systems.

NHK has proposed the MUSE (MUltiple Sub-Nyquist Encoding) system[19] for HDTV transmission. Taking the advantage of wide bandwidth satellite transmission, low data rate quincunx sampling technique and MAC format, it produces a high quality picture, with no crosstalk and high resolution in still picture.

CBS proposal[20] is to utilize two channels in MAC signal format. One channel transmits center 4:3 from 5:3 aspect picture with average adjacent two lines. The other channel transmits one field of 1050 interlaced lines as well as the "side panels" of the first channel. This is illustrated in Fig.3.6.

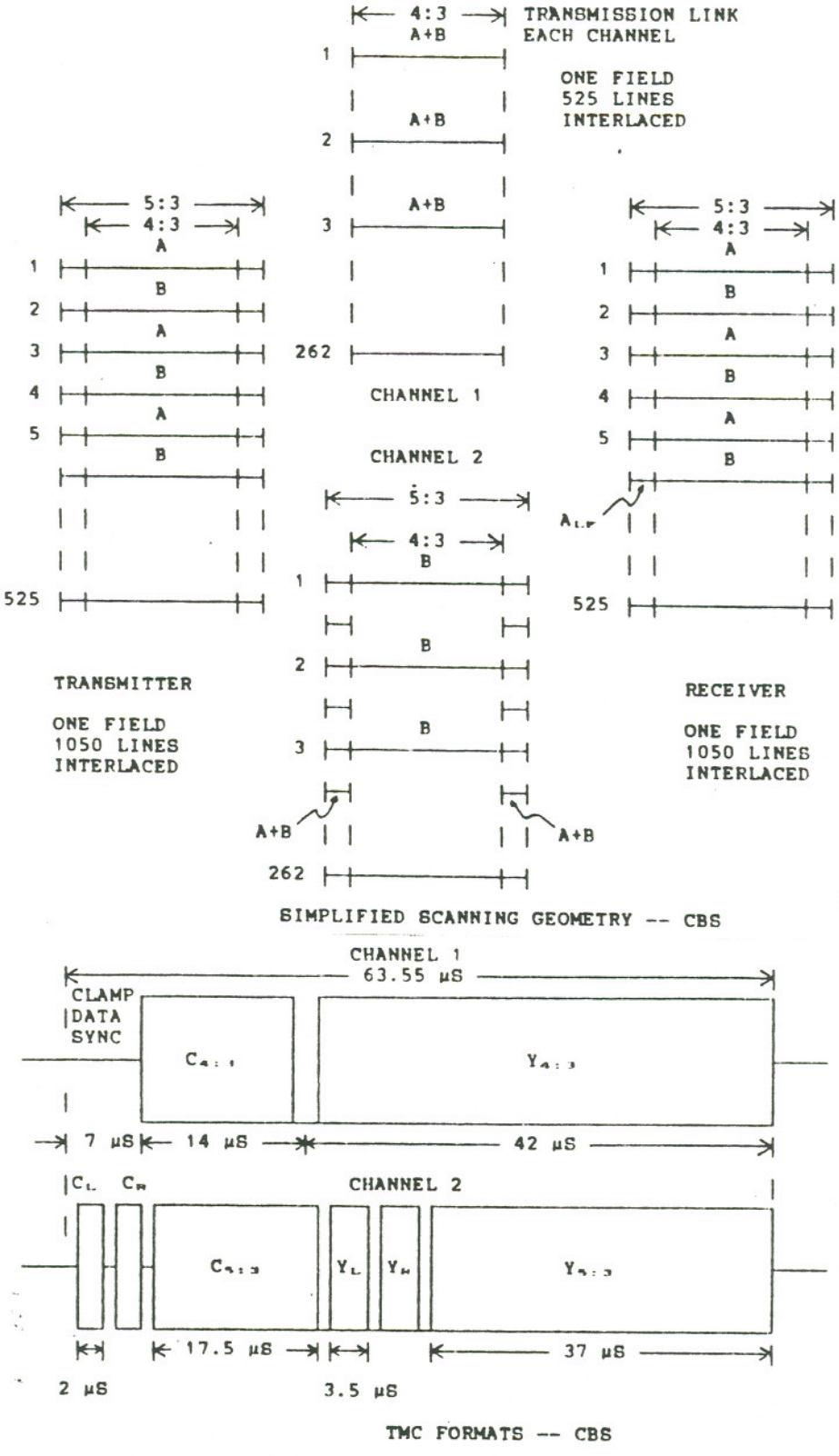


Fig.3.6 CBS proposal

Bell Laboratories proposal[21] presents a two channel system (see Fig.3.7). One channel transmits normal form NTSC signal with a vertical filter to convert 1050 lines (HDTV) into 525 line form. The other channel contains additional information, for instance, high frequency luminance and color difference information, two times horizontal resolution than that of normal NTSC, wide aspect ratio picture and multiple channel sound.

Glenn proposal[22] divides the signal into two groups, high temporal resolution information as motion picture and high spatial resolution information as still picture. For the former (motion portion), NTSC signal is used. For the later (still portion), MAC format and lower frame rate quincunx sampling are used.

Del Rey Group proposal[23] uses subpixel concept, which is to subdivide one pixel into smaller units, to increase both horizontal and vertical resolution. For example, TriScan as displayed in Fig.3.8 doubles the vertical resolution and triples the horizontal resolution and sacrifices the temporal resolution instead. So, Dual resolution processors ---- still and motion processor are required.

One of North American Philips proposal[24,25] is to use two channel transmission. The first channel is obtained by selecting a 4:3 aspect ratio portion from 16:9 aspect ratio picture of one interlaced field. The second channel carries side panels, progressive scan information and digital stereo sound. Both signals are normal NTSC signals. Fig.3.9 gives these two NTSC signals formation.

NBC has proposed a system[26] which uses single NTSC channel with pre-

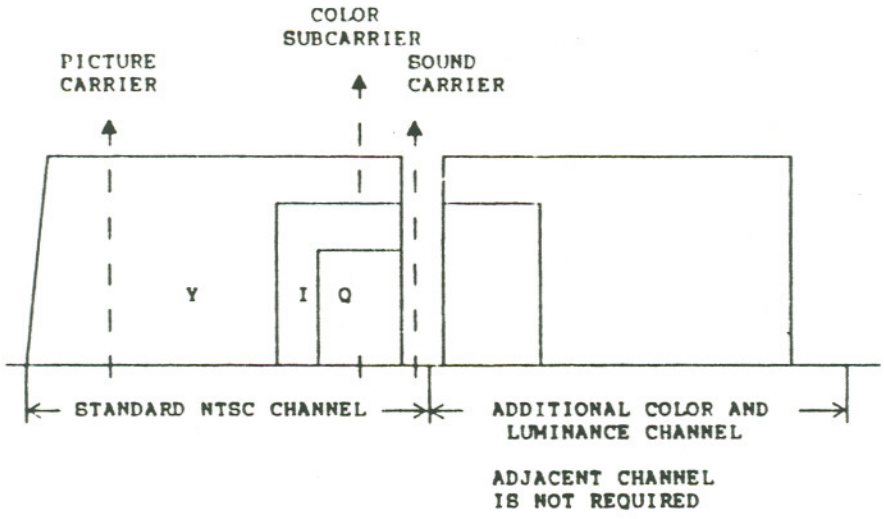


Fig.3.7 Bell labs proposal

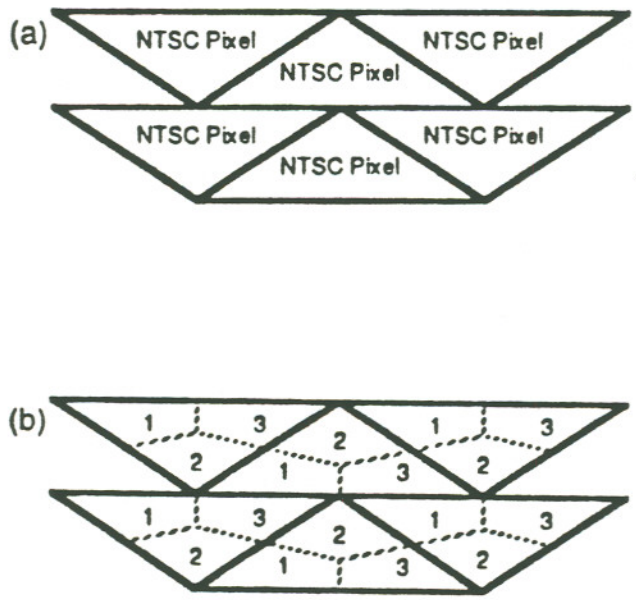


Fig.3.8 Concept of Tri Scan

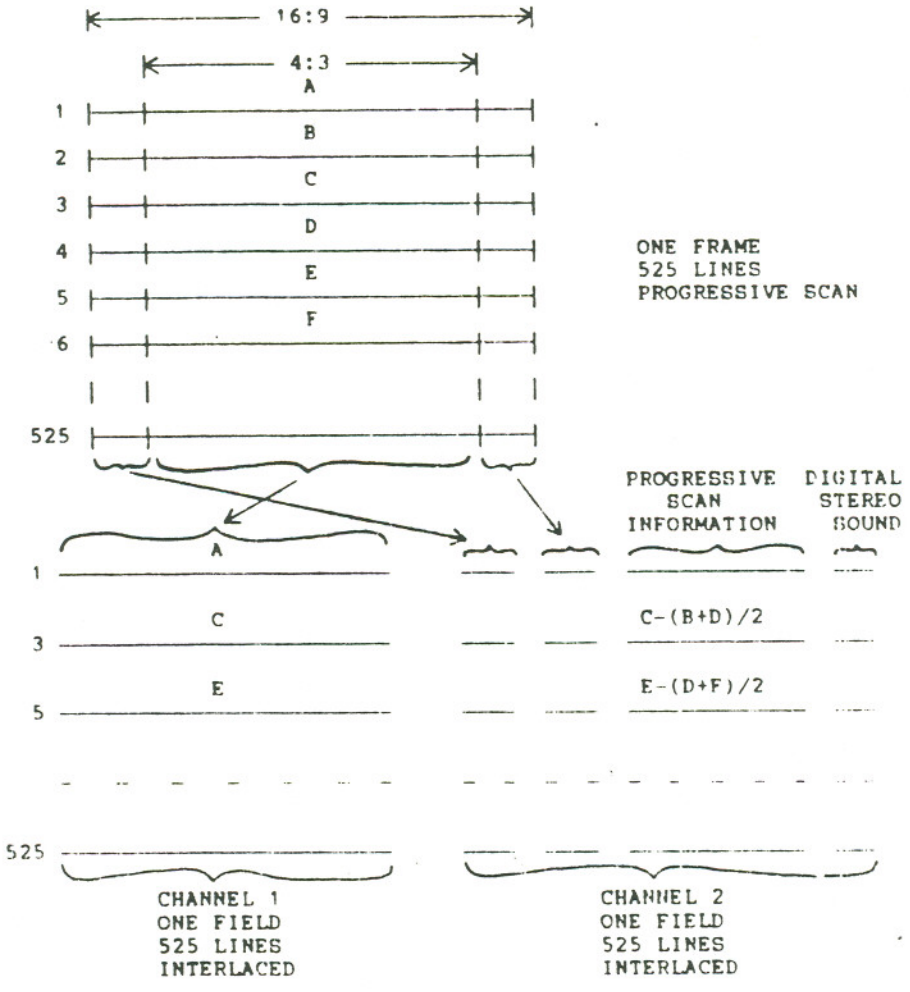


Fig.3.9 N.A. Philips proposal

combing filter, Fukinuki procedure, QUME modulation, higher line number scan and so on. A diagram of the spectrum of the transmitted signal is shown in Fig.3.10.

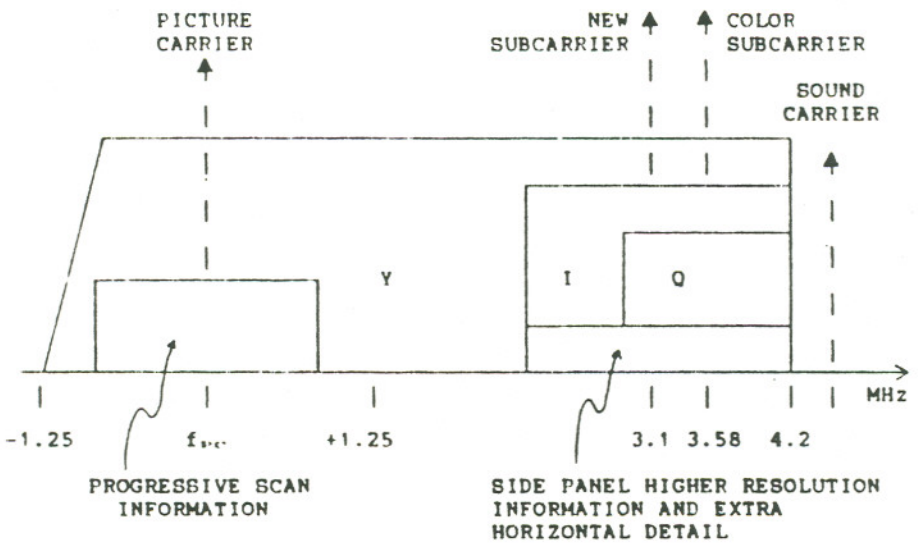


Fig.3.10 NBC proposal

INTRODUCTION OF THREE DIMENSIONAL FILTER

The goal of television signal processing is to obtain a good quality of picture. The television signal processing is in principle multidimensional filtering. The main purpose of a multidimensional filter is to limit the spectrum of the input signal as desired. In high definition video processing, 3-D filters are widely used, for example, to reduce the data rate, 3-D subsampling and filtering is used; to get high spatio-temporal resolution, deinterlacing filter and interpolating filter are applied; to control aliasing and crosstalk, pre- and post- filter are used in transmitter and receiver. However, it is neither realizable nor desirable to have an "ideal" filter with unit passband and zero elsewhere. Thus, the problem at hand is to design filters that perform the required spectral shaping of the signal without introducing undue degradations.

Distinguished from 1-D filter design[53], 3-D filter design is sophisticated since the multidimensional parameters can not be factored in general. Out of the various types of 3-D filters, we will deal only with linear shift time invariant filters.

Reviewing linear shift time invariant system, one important property is that the response of the filter to a sinusoidal signal is a sinusoid of the same frequency but with the amplitude and the phase modified. Another feature of 3-D filter, different from 1-D filter, is separability. A 3-D filter is said to be fully separable

if it can be expressed as the cascade of a one dimensional horizontal filter, a one dimensional vertical filter and a one dimensional temporal filter. The filter can also be partially separable, as a product of a one dimensional filter and a two dimensional filter. Separability reduces complexity in both filter design and filter implementation. However, the performance loss is not tolerable sometimes. Therefore, application conditions should be taken into account for separable filter selection.

For the various filters applied in TV signal processing, a common problem is that the effect of the filter is dependent upon the motion in the scene. For example, an attractive method for HDTV signal processing is skipping the frames at the transmitter to achieve low data rate and interpolating the skipped frames at the receiver to get high resolution. But if only linear frame interpolation filter is used, blurring will occur in the moving area and this degradation is visible in proportion to the speed of movement. Therefore, motion adaptive method is preferred. For this purpose, motion detection and estimation is the key point and it will be discussed in next chapter.

In this chapter, some specific purpose 3-D filters in HDTV are discussed based on motion state.

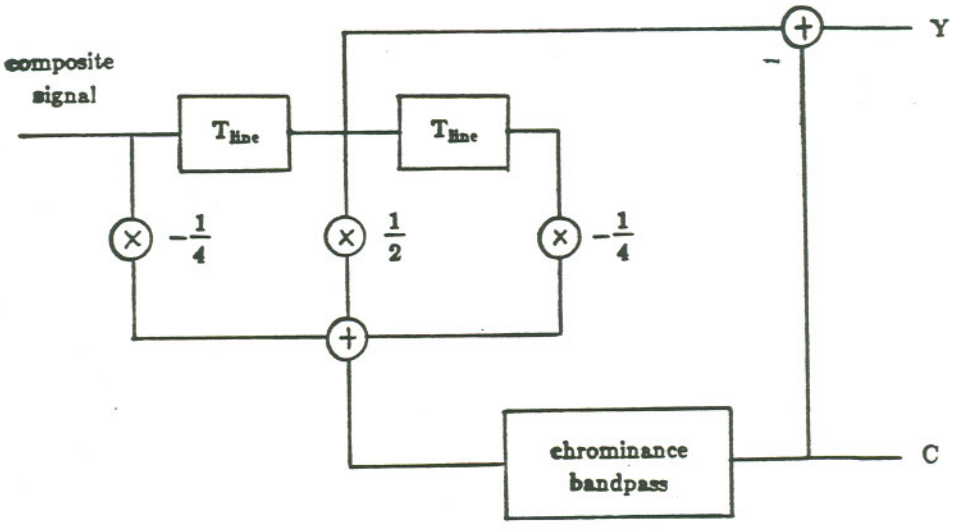
4.1. Pre- and Post- Filter for Crosstalk Elimination

Due to rather imperfect filtering, present NTSC TV picture suffers from crosstalk and loss of resolution. With TV system's characteristics, scanning

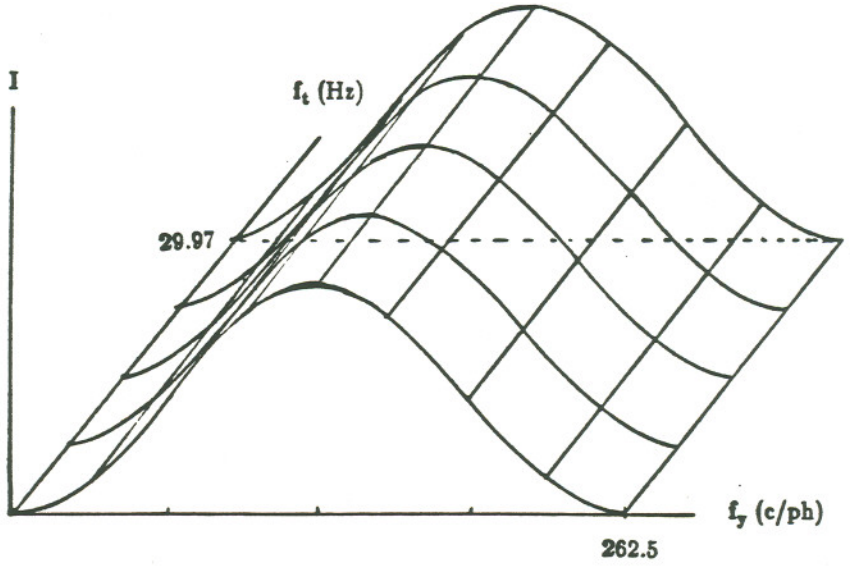
scheme presents the spectrum of TV signal (Fig.2.3) to have periodic components of frequencies f_{line}, f_{field} and f_{frame}. This gives the idea to separate luminance and chrominance signals by a periodic filter. This filter whose coefficients have a periodic frequency response resembling the shape of a comb, and is sometimes referred to as a comb filter[10,54]. Obviously, line comb filter, field comb filter and frame comb filter have a better matching to the TV signal as mentioned in Chapter 2. When they act in the system, cross color and cross luminance effects can be reduced and some amount of detail resolution can be preserved. In order to fully exploit the NTSC system capabilities, an adequate pre- comb filter is introduced in the encoder so that the pure luminance and the chrominance signal combine together to make up of a good composite signal. In contrast, an adequate post- comb filter is applied in the decoder to separate the luminance and the chrominance so as to eliminate the crosstalk. The simplest comb filter is one delay canceller, but it does not have a required shape. At least two delays are needed for cancellation. Moreover, for these three kinds of comb filters, it is obvious that there is no action along horizontal direction. Thus, vertical temporal two dimensional spectrum is more useful and convenient to interpret the characteristics of these filters.

4.1.1. Line Comb Filter

A line comb filter for NTSC signals requires a minimum of two lines delays. The filter characteristics is obtained from Eq.(4.1) and Eq.(4.2) with a sine or cosine magnitude response as shown in Fig.4.1. There is no change of the filter



(a) Block diagram of line comb filter



(b) Transfer function of line comb filter

Fig.4.1 Scheme and spectrum of line comb filter

frequency response both in horizontal and temporal direction and two dimensional description of chrominance response is shown in Fig.4.1(b) as well. Luminance response is nothing but $1-H_c(f_x,f_y,f_t)$ given by Eq.(4.3).

$$r_c(x,y,t) = -\frac{1}{4}\, i(x,y+y_l,t) + \frac{1}{2}\, i(x,y,t) - \frac{1}{4}\, i(x,y-y_l,t) \tag{4.1}$$

where $\quad y_l = \dfrac{1}{f_y} = \dfrac{1}{\frac{525}{2}}$ picture height

$$R_c(f_x,f_y,f_t) = [-\frac{1}{4}\, e^{j2\pi y_l f_y} + \frac{1}{2} - \frac{1}{4}\, e^{-j2\pi y_l f_y}]\; I(f_x,f_y,f_t)$$

$$= \frac{1}{2}\,[1 - \cos(2\pi y_l f_y)]\; I(f_x,f_y,f_t)$$

$$= H_c(f_x,f_y,f_t)\; I(f_x,f_y,f_t)$$

$$H_c(f_x,f_y,f_t) = \frac{1}{2}\,[1 - \cos(2\pi y_l f_y)] \tag{4.2}$$

$$H_l(f_x,f_y,f_t) = 1 - H_c(f_x,f_y,f_t) = \frac{1}{2}\,[1 + \cos(2\pi y_l f_y)] \tag{4.3}$$

Because of no degradation in temporal direction, the line comb filter is more powerful for separating luminance and chrominance signals from moving pictures. Since the filter is not ideal --- it has neither unit passband, nor zero stopband, nor short transition band. There still remains small amount of crosstalk for certain spatio-temporal frequencies.

4.1.2. Frame Comb Filter

With the same structure as line comb filter (Fig.4.1(a)), the frame comb filter (Fig.4.2(a)) is different only in the delay time --- the T_{frame} replaces T_{line} in temporal direction. The orientation of the chrominance filter performance thereby changes from vertically to temporally as illustrated in Fig.4.2(b). The effect of this frame comb filter can be derived from the frequency characteristics.

$$r_c(x,y,t) = -\frac{1}{4}\, i(x,y,t+t_f) + \frac{1}{2}\, i(x,y,t) - \frac{1}{4}\, i(x,y,t-t_f) \tag{4.4}$$

where $\quad t_f = \dfrac{1}{f_{frame}} = \dfrac{1}{30}$ seconds

$$R_c(f_x,f_y,f_t) = [-\frac{1}{4}\, e^{j2\pi t_f f_t} + \frac{1}{2} - \frac{1}{4}\, e^{-j2\pi t_f f_t}]\; I(f_x,f_y,f_t)$$

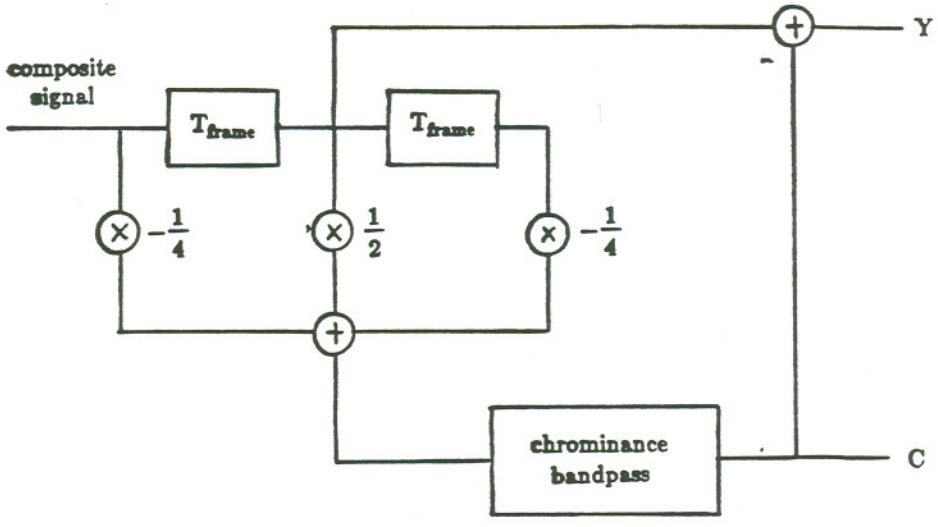
$$= \frac{1}{2}\,[1 - \cos(2\pi t_f f_t)]\; I(f_x,f_y,f_t)$$

$$= H_c(f_x,f_y,f_t)\; I(f_x,f_y,f_t)$$

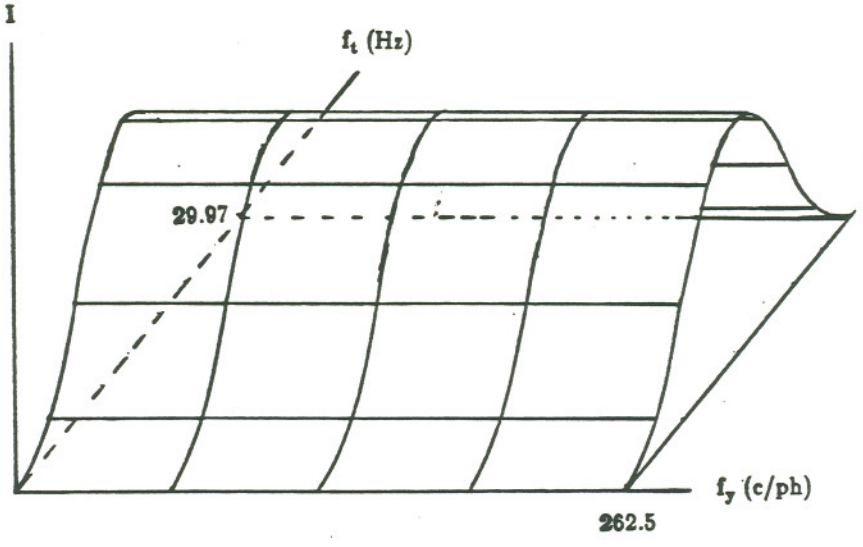
$$H_c(f_x,f_y,f_t) = \frac{1}{2}\,[1 - \cos(2\pi t_f f_t)] \tag{4.5}$$

$$H_l(f_x,f_y,f_t) = 1 - H_c(f_x,f_y,f_t) = \frac{1}{2}\,[1 + \cos(2\pi t_f f_t)] \tag{4.6}$$

From Fig.4.2(b), luminance is attenuated in the vicinity of 15 Hz temporal frequency. This means that for all stationary images, there is no cross color and cross luminance effects and full detail spatial resolution will be maintained.



(a) Block diagram of frame comb filter



(b) Transfer function of frame comb filter

Fig.4.2 Scheme and spectrum of frame comb filter

While for moving pictures, the detail information will be lost. For some spatio-temporal frequencies, crosstalk effects accompany the signal again as this is not an "ideal" filter.

4.1.3. Field Comb Filter

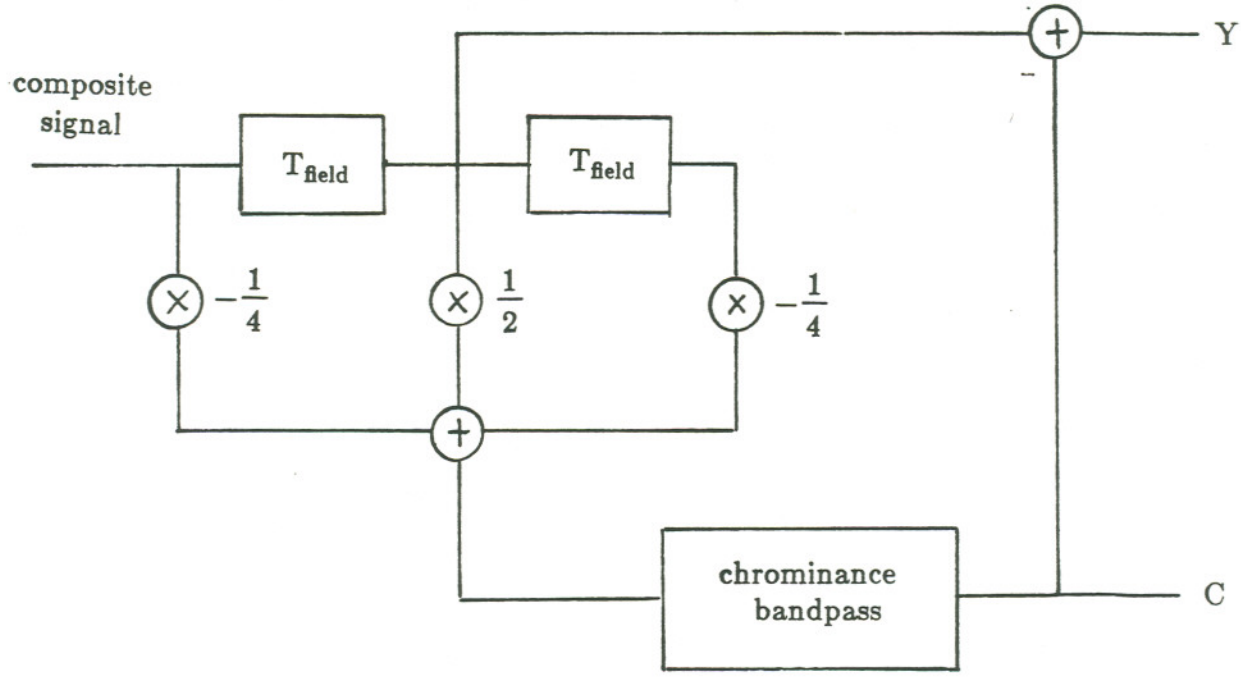
Compromise between the line comb filter and the frame comb filter, a filter suitable for both stationary and moving image is the field comb filter. With the benefits of smaller delay space than frame comb filter, the temporal resolution is increased. There are two types of field comb filters. One chooses the line above in the previous field, corresponding to a delay of 262 lines. The other picks up the lines below in the previous field, corresponding to 263 lines. These filters have half line delay in vertical and one field in temporal direction respectively. The frequency response will be deduced by Eq.(4.7) and Eq.(4.8). The filter structure and its frequency response are given by Fig.4.3.

$$r_c(x,y,t) = -\frac{1}{4}i(x,y\mp y_{f'},t+t_{f'}) + \frac{1}{2}i(x,y,t) - \frac{1}{4}i(x,y\pm y_{f'},t-t_{f'}) \qquad (4.7)$$

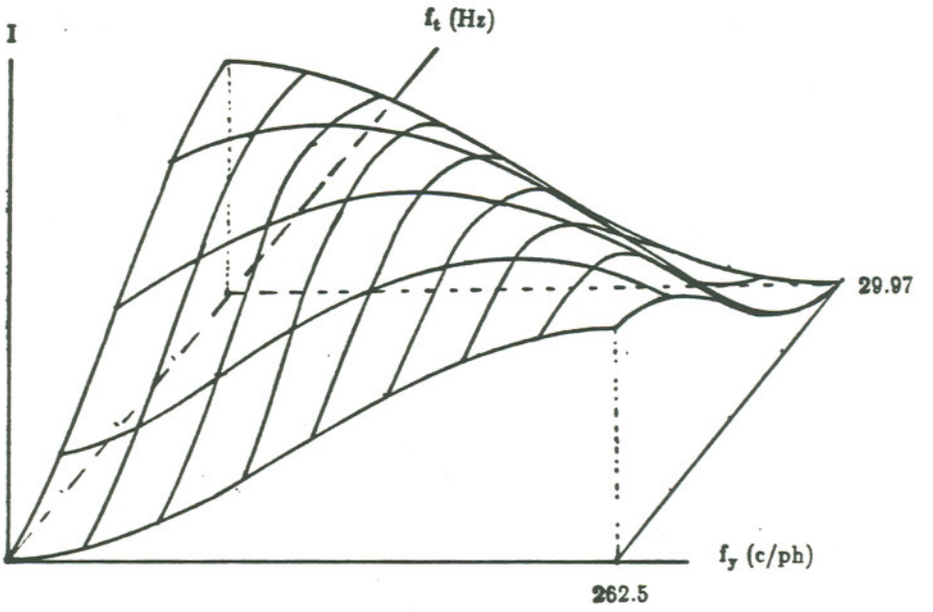
where $y_{f'} = \dfrac{0.5}{f_y} = \dfrac{0.5}{\frac{525}{2}}$ picture height

$$t_{f'} = \frac{1}{f_{field}} = \frac{1}{60} \text{ seconds}$$

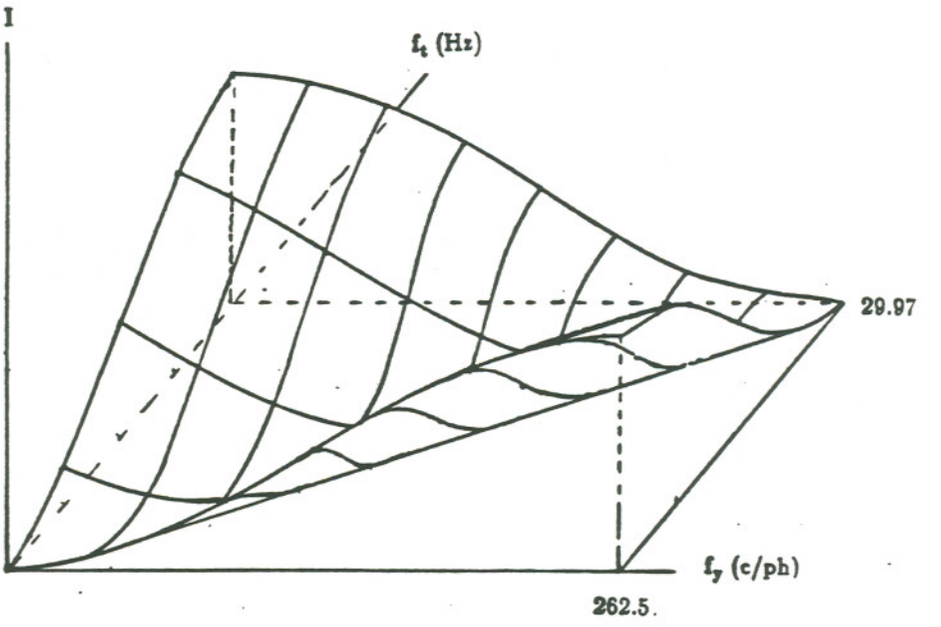
$$R_c(f_x,f_y,f_t) = [-\frac{1}{4}e^{j2\pi(\mp y_{f'}f_y + t_{f'}f_t)} + \frac{1}{2} - \frac{1}{4}e^{j2\pi(\pm y_{f'}f_y - t_{f'}f_t)}]\ I(f_x,f_y,f_t)$$



(a) Block diagram of field comb filter



(b) Transfer function of field comb filter (262 lines delay)



(c) Transfer function of field comb filter (263 lines delay)

Fig.4.3 Scheme and spectrum of field comb filter

$$= \frac{1}{2}\,[1 - \cos(2\pi(\mp y_{f'}f_y + t_{f'}f_t))]\; I(f_x,f_y,f_t)$$

$$= H_c(f_x,f_y,f_t)\; I(f_x,f_y,f_t)$$

$$H_c(f_x,f_y,f_t) = \frac{1}{2}\,[1 - \cos(2\pi(\mp y_{f'}f_y + t_{f'}f_t))] \tag{4.8}$$

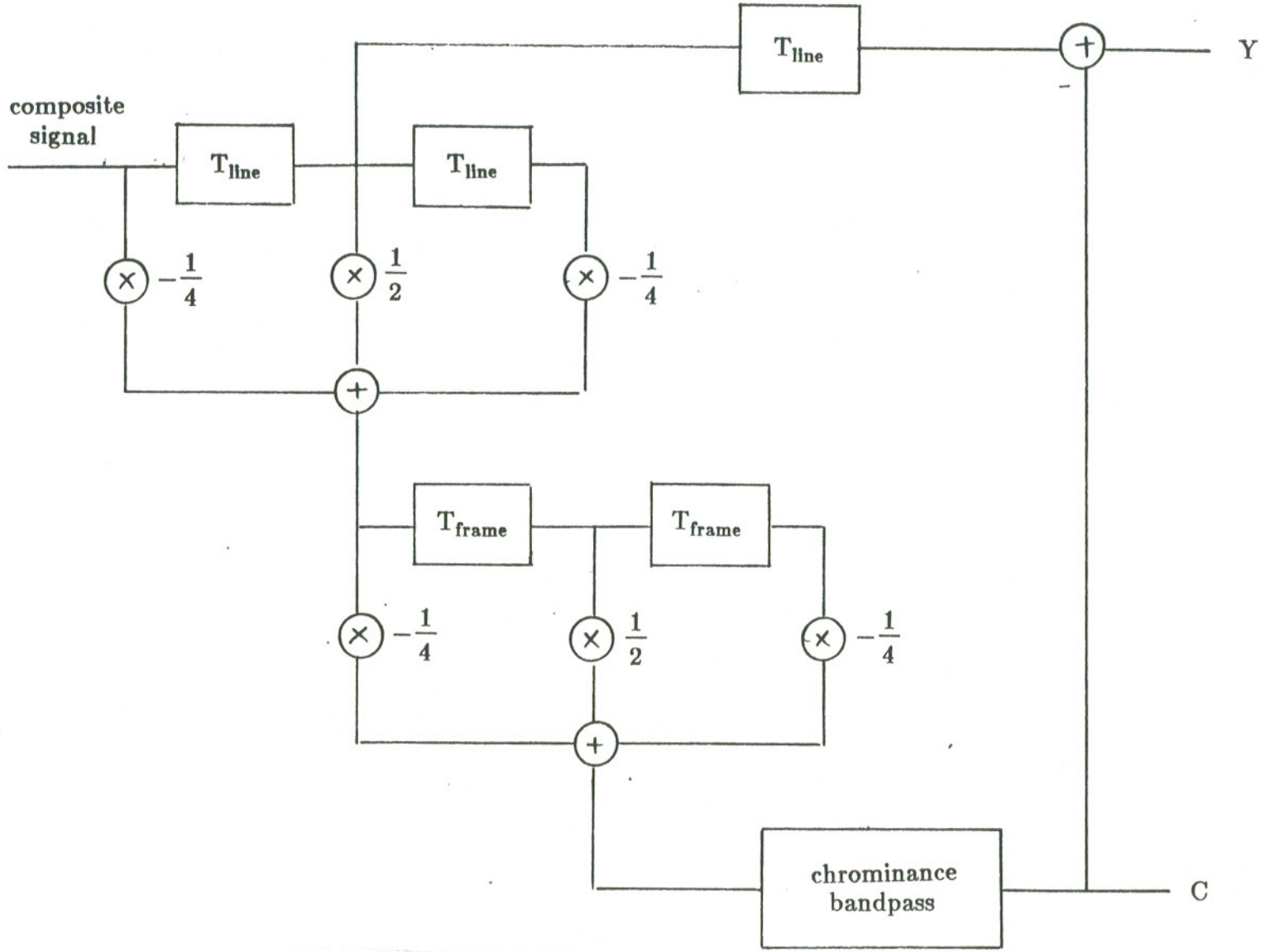
$$H_l(f_x,f_y,f_t) = 1 - H_c(f_x,f_y,f_t) = \frac{1}{2}\,[1 + \cos(2\pi(\mp y_{f'}f_y + t_{f'}f_t))] \tag{4.9}$$

4.1.4. Combination of Basic Comb filters

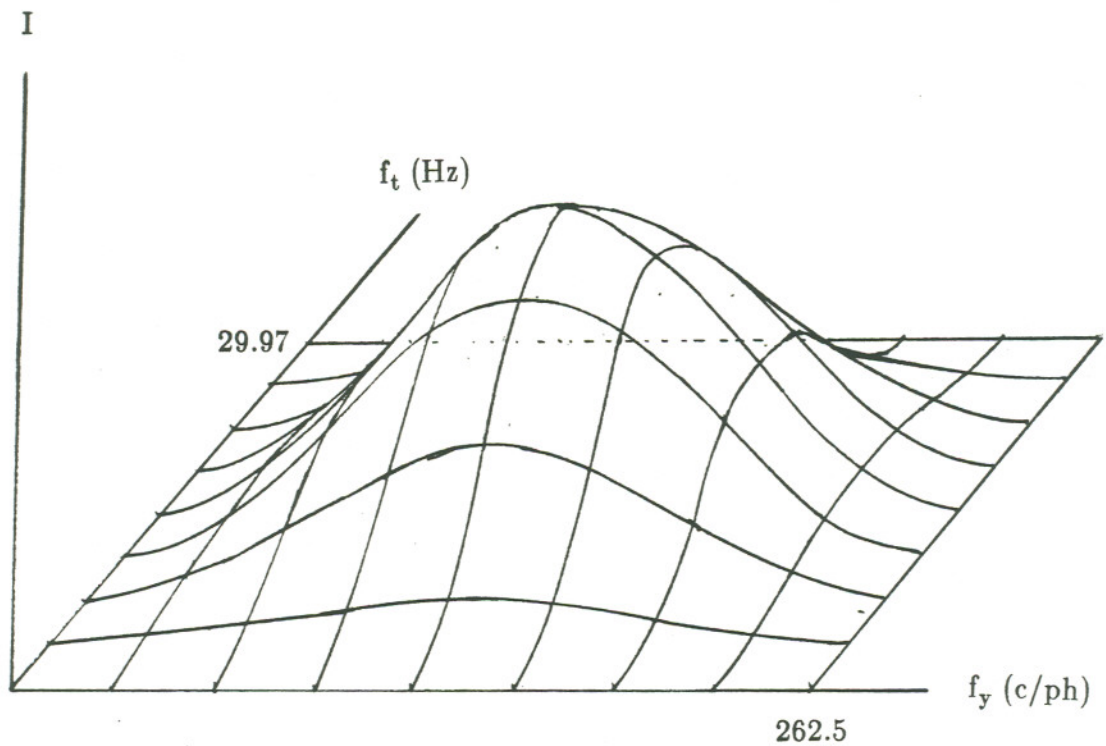
As analyzed above, the basic comb filters have their own available and efficient circumstances. To take advantage of single filter attitudes, an adaptive mode can be realized by combining them in series or parallel structure. For example, if a line comb filter connects with a frame filter in series Fig.4.4, the resolution will be ameliorated both in stationary and moving picture. However, the chrominance resolution is sacrificed. By means of a motion detection, more suitable filter can be selected adaptively to match different conditions. On the other hand, an adaptive filter with feedback coefficient control[45] can be used to change the shape of the filter to eliminate the crosstalk. But stability etc. should be considered carefully.

4.2. Filter for Low Data Rate Transmission

As mentioned in Chapter 3, in order to decrease data rate, various sampling



(a) Block diagram of frame and line series combined comb filter



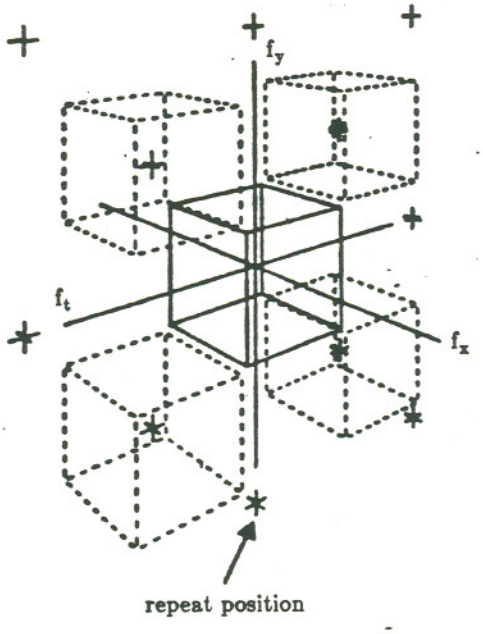
(b) Spectrum of diagram of frame and line series combined comb filter

schemes are implemented. The frequency response of signal will consist of the original signal and the replicas after certain frequency shifts (Eq.(4.10)).

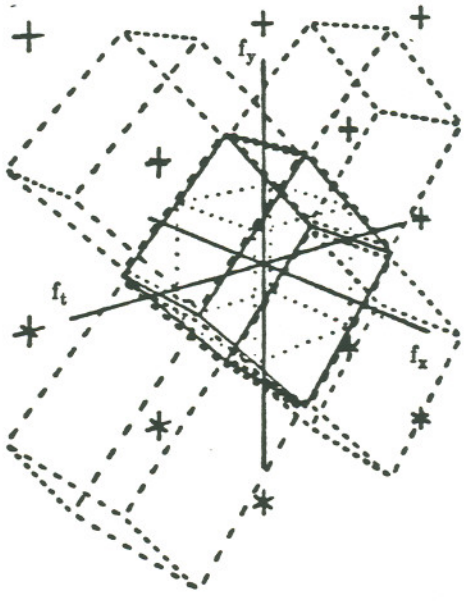
$$I_s(f_x,f_y,f_t) = \frac{1}{T_{xs}T_{ys}T_{ts}} \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} I(f_x - lf_{xs}, f_y - mf_{ys}, f_t - nf_{ts}) \quad (4.10)$$

To avoid aliasing, a lowpass filter[28] applied on basic signal makes the signal spectrum confined in the box of $\frac{f_{xs}}{2}$ wide, $\frac{f_{ys}}{2}$ long and $\frac{f_{ts}}{2}$ deep. Consequently, the cost is missing high resolution either in horizontal or in vertical or in temporal direction. This defect can be overcome by an adaptive use of sampling filters. We know, the eye is somewhat less sensitive to high spatial frequencies if they are moving than if they are stationary. So, for our purpose, we can adapt sampling in still picture and select wide band of sampling or no sampling in moving picture. Of course, motion detection is required.

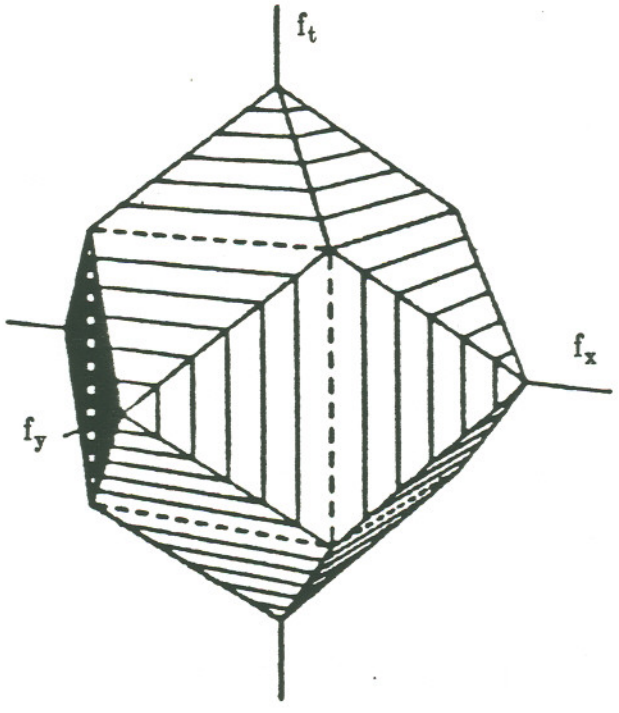
It is worth noticing the efficient bandwidth of the sampling filter. For example[27], Fig.4.5(a) is a perspective view of a cubic bandwidth. Since the vertical resolution is sometimes unnecessarily high when the temporal frequency is high, the marginal bandwidth to avoid aliasing could be an orthorhombic. The space then is filled up (Fig.4.5(b)). With the same idea, the horizontal resolution can be sacrificed when the vertical and temporal frequencies are high. This gives the dodecahedron bandwidth (Fig.4.5(c)). As the knowledge of absolutely optimal bandwidth does not exist, a trade off manipulation can be chosen under the available and feasible condition.



(a) Conventional television spectrum



(b) Efficient television spectrum



(c) One optional dodecahedron spectrum

Fig.4.5 Spectrum of sampling

4.3. Filter for High Resolution Display

High resolution display includes high spatial resolution display and high temporal resolution display. The principal method for high resolution display is interpolation[55,56] with the aid of transmitted data.

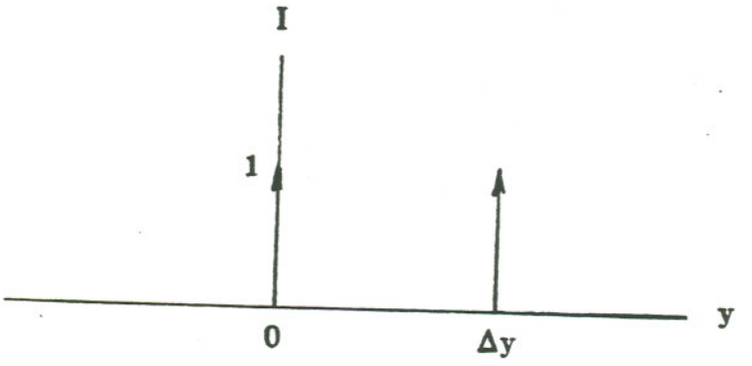
4.3.1. High Vertical Resolution Display (Deinterlacing)

One kind of high resolution display is the filter for interlaced scanning transferred to progressive scanning. It is a type of spatial filter to increase vertical resolution. As introduced in Chapter 3, the common procedure is line repeat or line average. It is reasonable to assume the video signal is continuous in horizontal direction and discrete in vertical and temporal directions.

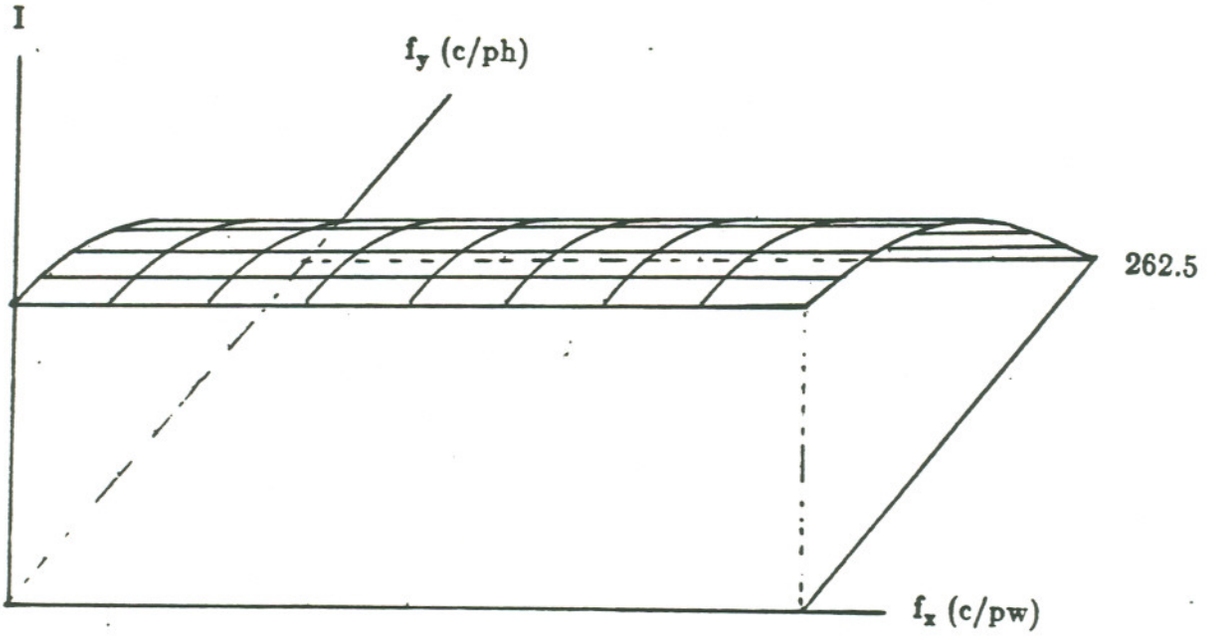
Consider the procedure of line repeat scheme. Since the output can be described as the convolution of the input signal and the impulse signal as shown in Fig.4.6(a), the impulse response of the filter is given by Eq.(4.11) and plotted in Fig.4.6(b).

$$\begin{aligned} H(f_x,f_y,f_t) &= \mathcal{F}\,[\delta(y) + \delta(y-\Delta y)] \\ &= 1 + e^{j2\pi\Delta y f_y} \\ &= 2\cos(\pi\Delta y f_y)\, e^{j\pi\Delta y f_y} \end{aligned} \tag{4.11}$$

where $\Delta y = \frac{1}{525}$ picture height



(a) Impulse response of line repeat deinterlacing filter



(b) Transfer function of line repeat deinterlacing filter

Fig.4.6 Spectrum of line repeat deinterlaced filter

For the line average scheme, the output can be considered as the convolution of the input signal and the impulse signal as shown in Fig.4.7(a). Accordingly, the impulse response of the filter is obtained by Eq.(4.12) and plotted in Fig.4.7(b).

$$H(f_x,f_y,f_t) = \mathcal{F}\,[\frac{1}{2}\,\delta(y+\Delta y) + \delta(y) + \frac{1}{2}\,\delta(y-\Delta y)]$$

$$= \frac{1}{2}\,e^{j2\pi\Delta y f_y} + 1 + \frac{1}{2}\,e^{-j2\pi\Delta y f_y}$$

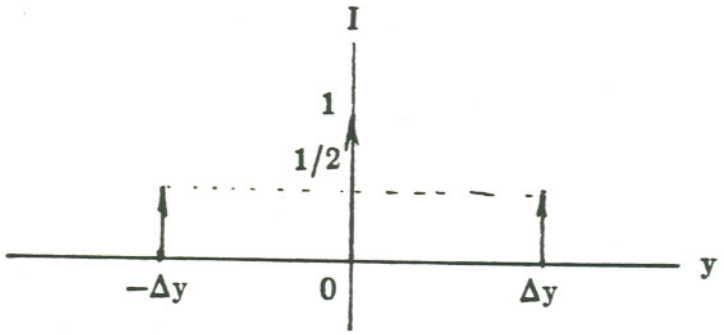
$$= 2\,\cos^2(\pi\Delta y f_{y)} \qquad (4.12)$$

where $\Delta y = \frac{1}{525}$ picture height

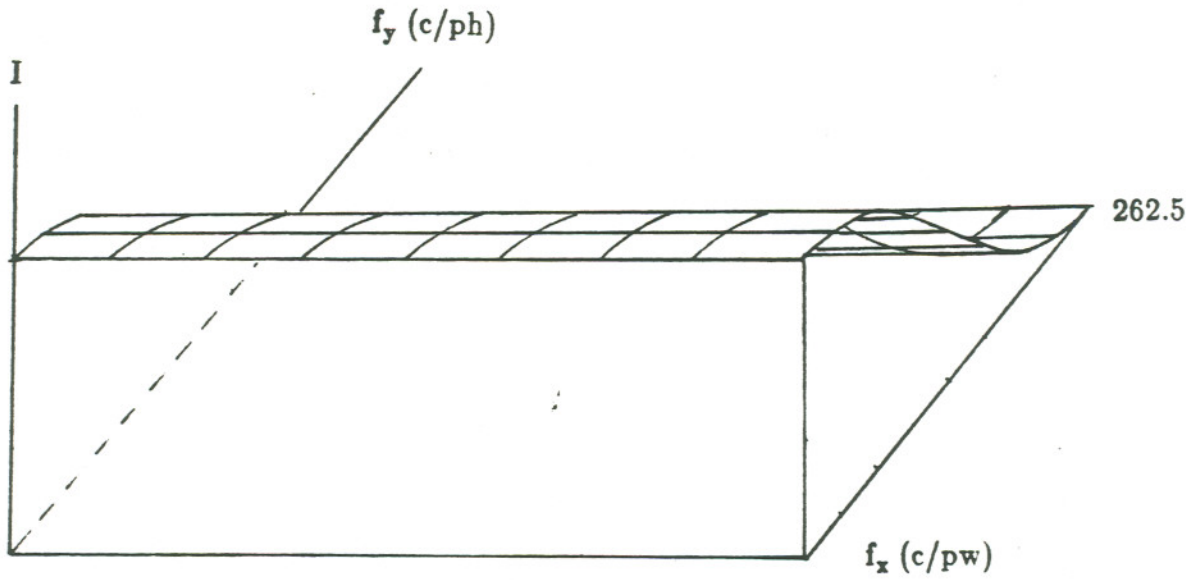
In general, interpolation signal with coefficients γ and $1-\gamma$ can be generated by convolution of the input signal and a sampled triangle signal. In our case, the triangle signal is plotted in Fig.4.8. whose frequency response corresponds to $2\Delta y Sa^2(2\pi f_y \Delta y)$. The sampled triangle signal is a phase shifted (y_0) signal, where y_0 depends on γ. When it is sampled at the rate of $2\Delta y$, its frequency response repeats with the period $\frac{1}{2\Delta y}$ and becomes $\sum_{l} Sa^2[2\pi\Delta y(f_y - \frac{l}{2\Delta y})]\,e^{-j2\pi f_y y_0}$. This results the spectrum of interpolated signal to be

$$F\,[i(x,y,t)] = \mathcal{F}\,[f(x,y,t) * (1 + tr(x,y,t))]$$

$$= F(f_x,f_y,f_t)\,[1 + Tr(f_x,f_y,f_t)]$$



(a) Impulse response of line average deinterlacing filter



(b) Transfer function of line average deinterlacing filter

Fig.4.7 Spectrum of line average deinterlaced filter

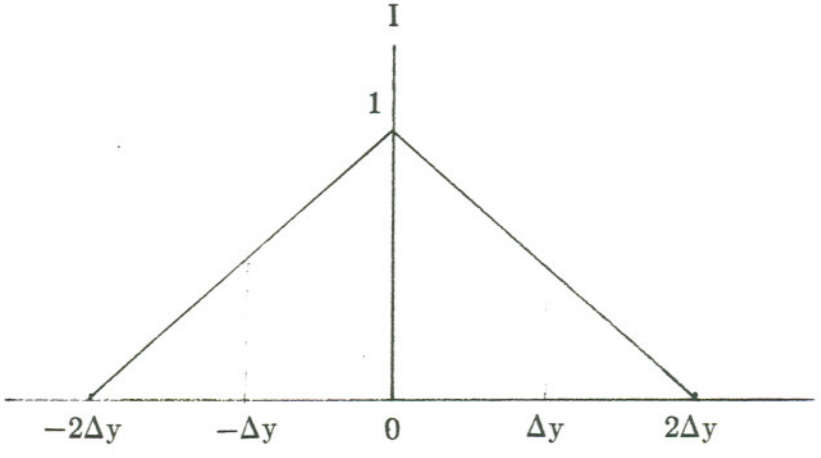


Fig.4.8 Triangle impulse response of interpolation

$$= F(f_x,f_y,f_t)\ [1 + \sum_{l} Sa^2(2\pi\Delta y(f_y - \frac{l}{2\Delta y}))\ e^{-j2\pi f_y y_0}] \qquad (4.13)$$

Therefore, the impulse response of line deinterlacing filter is obtained by

$$H(f_x,f_y,f_t) = 1 + \sum_{l} Sa^2[2\pi\Delta y(f_y - \frac{l}{2\Delta y})]\ e^{-j2\pi f_y y_0} \qquad (4.14)$$

In the case of $\gamma = \frac{1}{2}$, $y_0 = \Delta y$, the impulse response of line average deinterlacing filter is

$$H(f_x,f_y,f_t) = 1 + \sum_{l} Sa^2[2\pi\Delta y(f_y - \frac{l}{2\Delta y})]\ e^{-j2\pi f_y \Delta y}$$

$$= 2 \sum_{k} Sa^2[2\pi\Delta y(f_y - \frac{k}{\Delta y})] \qquad (4.15)$$

It can also be considered as the convolution of the input signal and the triangle signal sampled with the rate of Δy.

Since they act in one field, there is no change in temporal direction, they can be used in either stationary picture or moving picture. In addition, horizontal and vertical two dimensional spectrum can depict the filter characteristics perfectly. From line repeat procedure (Fig.4.6), although high resolutions attenuate in both vertical and diagonal direction, the high spatial frequencies are still so high to generate a sharp picture. The oblique line distortion is such an example. From line average procedure (Fig.4.7), high frequencies attenuate so quickly in both vertical and diagonal direction that the picture is smoothed and the aliasing

is improved.

4.3.2. High Temporal Resolution Display

Flicker elimination filter is an another kind of high resolution display filter. It is a temporal filter. As described in Chapter 3, there are three common basic flicker elimination filters: temporal average filter, field repeat filter and picture repeat filter. For simplicity, here we discuss only the first two filters. As above, the frequency response of these filters can be derived from Eq.(4.16) and Eq.(4.17) for first two filters respectively, but the spectra now are oriented in temporal direction and are described by vertical temporal two dimensions (Fig.4.9 and Fig.4.10).

$$H(f_x,f_y,f_t) = \mathcal{F}\,[\frac{1}{2}\,\delta(t+\frac{\Delta t}{2}) + \delta(t) + \frac{1}{2}\,\delta(t-\frac{\Delta t}{2})]$$

$$= \frac{1}{2}\,e^{j\pi\Delta t f_t} + 1 + \frac{1}{2}\,e^{-j\pi\Delta t f_t}$$

$$= 2\,\cos^2\frac{\pi\Delta t f_t}{2} \qquad (4.16)$$

where $\Delta t = \frac{1}{60}$ seconds

$$H(f_x,f_y,f_t) = \mathcal{F}\,[\delta(t) + \delta(t-\frac{\Delta t}{2})]$$

$$= 1 + e^{-j\pi\Delta t f_t}$$

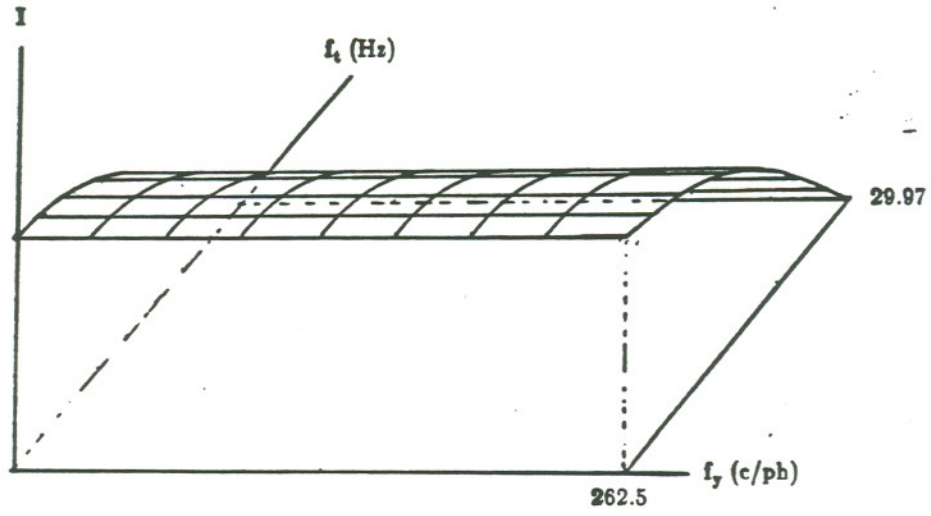


Fig.4.9 Transfer function of field repeat filter

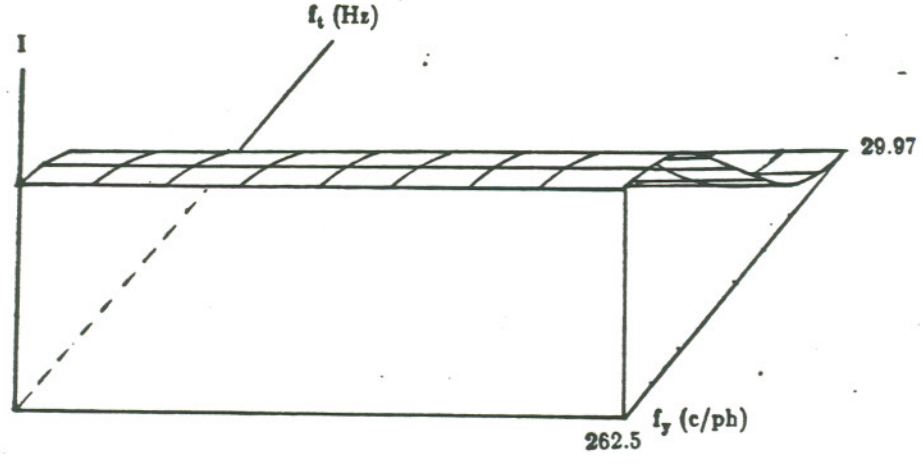


Fig.4.10 Transfer function of temporal average filter

$$= 2\cos\frac{\pi\Delta t f_t}{2}\, e^{-j\frac{\pi\Delta t f_t}{2}} \tag{4.17}$$

where $\Delta t = \frac{1}{60}$ seconds

The frequency response of all these filters decays at high temporal frequency, i.e. in moving picture. That is why the edges get blurred in moving areas after filtering. If high spatial resolution is still required in the moving picture, motion compensation has to be provided.

The simpliest motion compensation is linear interpolation in three dimensions. Same as the above, the spectrum is illustrated by Eq.(4.18).

$$H(f_x,f_y,f_t) = \mathcal{F}[\frac{1}{2}\,\delta(x+\frac{\Delta x}{2},y+\frac{\Delta y}{2},t+\frac{\Delta t}{2})+\delta(x,y,t)+\frac{1}{2}\delta(x-\frac{\Delta x}{2},y-\frac{\Delta y}{2},t-\frac{\Delta t}{2})]$$

$$= \frac{1}{2}\, e^{-j\pi(\Delta x f_x+\Delta y f_y+\Delta t f_t)} + 1 + \frac{1}{2}\, e^{-j\pi(\Delta x f_x+\Delta y f_y+\Delta t f_t)}$$

$$= 2\cos^2[\frac{\pi}{2}(\Delta x f_x+\Delta y f_y+\Delta t f_t)] \tag{4.18}$$

where $\Delta t = \frac{1}{60}$ seconds

Δx and Δy determine the motion vector.

After motion compensation, the picture is smoothed in spatial temporal direction and the picture quality therefore does not degrade.

MOTION ESTIMATION

Motion considerations have become more and more important in TV signal processing. One of the main developments is the application of mathematical models describing the motion of objects. Because of complex computation and real-time processing requirement, the motion models are commonly based on rigid body assumption. A simple model is a translational one. Rotation model is also introduced to match the motion closely. A translation movement generates a displacement of the moving body between frames. A displacement vector indicates the motion direction and the motion velocity as well. A rotation movement brings a spin with a certain angle and with respect to a certain axis between frames. Knowing these vectors, angles and axes, we can use them to realize motion compensation. So, the key point is to find out these parameters, i.e. motion estimation. Essentially, the estimation algorithms[57–81] can roughly be classified into block matching algorithms[57], recursive algorithms[57] and phase correlation algorithms[57], feature based motion estimation algorithms[58] and optic flow based motion estimation algorithms[58]. Another important branch is motion detection. Much less complex than motion estimation, it only needs to roughly appraise what speed region the motion is in, for instance, to determine whether scenes have slow movement or medium movement or fast movement. Therefore, we can choose different sampling patterns, apply different processing, etc. Suppose we divide frequency spectrum as blocks shown in Fig.5.1, where

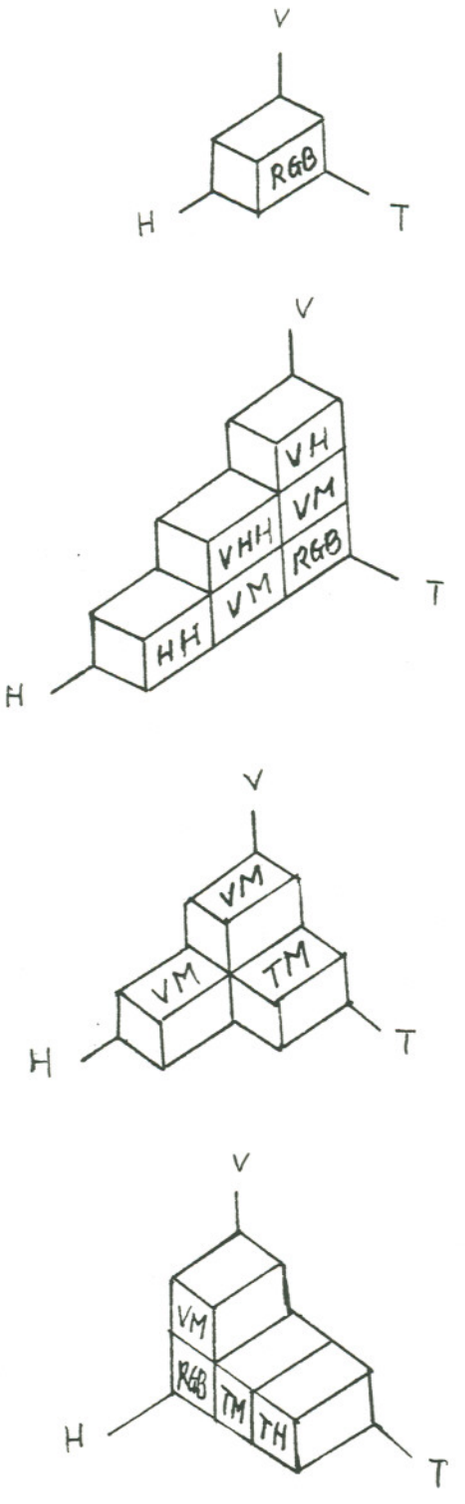


Fig.5.1 Signal transmitted by motion detection

RGB stands for basic red, green and blue band; VH, VHH, HH and TH are shorted for high vertical frequencies, high diagonal frequencies, high horizontal frequencies and high temporal frequencies, respectively; VM, HM and TM are abbreviations of medium vertical frequencies, medium horizontal frequencies and medium temporal frequencies, respectively. Apparently, basic band signal RGB must be transmitted all the time. When the motion is slow, high spatial resolution is required based on human visual system, then VH, VHH, HH, VM, HM and RGB have to be sent out. When the motion is fast, human eyes are not so sensitive to the spatial resolution, then VM, TM, TH and RGB are enough. When the motion is medium, VM, HM, TM and RGB should be arranged so as to meet the needs in both spatial and temporal requirement.

5.1. Basic Algorithms for Estimation of Displacement

A moving object in a scene generates frame-to-frame luminance changes. These luminance variations can be used to estimate the parameters of a mathematical model describing the motion of objects. For example, the displacement D for a simple moving edge as shown in Fig.5.2 can be estimated by

$$\hat{D} = d\hat{x} = \frac{\sum_{M} |FD|}{\sum_{M} |ED|} \tag{5.1}$$

where |FD| represents the magnitude of the frame difference signal, |ED| denotes the magnitude of the element difference signal and M is the area which is defined by frame differences greater than a given threshold. By the way, this equation

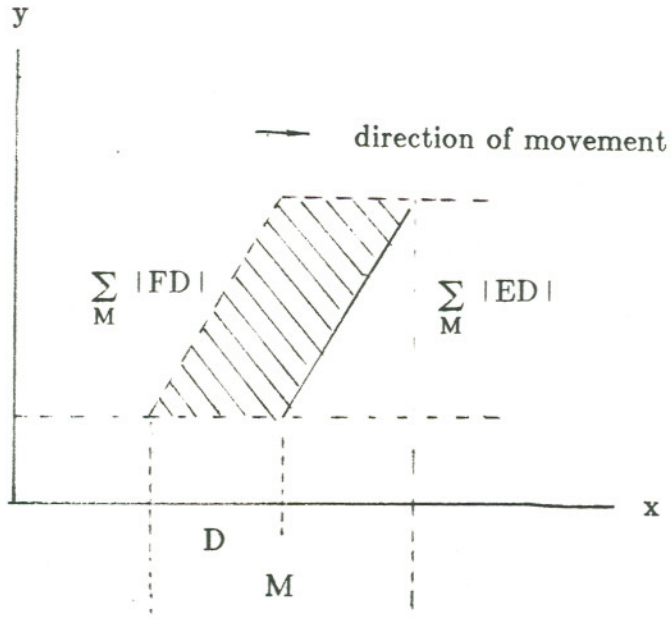


Fig.5.2 Illustration of the displacement estimation scheme

does not indicate the direction of the displacement. For small displacements D with components dx and dy, the frame differences FD can be generated as

$$\begin{aligned} FD(x,y) &= i_k(x,y) - i_{k-1}(x,y) \\ &= i_k(x,y) - i_k(x+dx,y+dy) \\ &= -\frac{\partial i_k(x,y)}{\partial x}dx - \frac{\partial i_k(x,y)}{\partial y}dy - n(x,y) \\ &\approx -D^T \cdot \nabla i_k(x,y) \end{aligned} \tag{5.2}$$

where D^T is transpose of D; $i_k(x,y)$ is the luminance value at point (x,y) in frame k; n(x,y) represents the higher order terms of the Taylor series expansion which can be neglected.

Considering a special case, the object moves only in x-direction. The estimate of the displacement vector $\hat{D}$ is given by

$$\hat{D} = d\hat{x} = -\frac{FD(x,y)}{\frac{\partial i(x,y)}{\partial x}} \tag{5.3}$$

If the boundary of the moving target is known, the statistical evaluation can be used to calculate the displacement over the entire area M, i.e.

$$d\hat{x} = -\frac{E[FD(x,y)\cdot\frac{\partial}{\partial x}i(x,y)]}{E[(\frac{\partial}{\partial x}i(x,y))^2]} = -\frac{\sum_M (FD\cdot ED)}{\sum_M (ED)^2} \tag{5.4a}$$

$$d\hat{y} = -\frac{E[FD(x,y)\cdot\frac{\partial}{\partial y}i(x,y)]}{E[(\frac{\partial}{\partial y}i(x,y))^2]} = -\frac{\sum_M(FD\cdot LD)}{\sum_M(LD)^2} \tag{5.4b}$$

in which LD is line difference.

This estimation is simple but it has to be confined in measuring small displacements. To overcome this problem, following methods have been developed.

5.1.1. Block matching estimation algorithms

A small portion of a moving object can be approximated to be displaced with the same speed. This gives the idea of block matching, which is nothing but to segment an image into a certain number of rectangular blocks, to calculate only one displacement vector for each block with optimal searching and place it in the center of the block. Obviously, it offers reducing the computation complexity.

Assume a block consists of M×N pixels centered at point (x,y) in frame k and moves possiblely dm pixels to the point x±dm , y±dm in frame k-1. Then, the searching area SA will be SA = (M+2dm)×(N+2dm). A simple way of the matching criterion is to evaluate the mean square error MSE or the mean absolute difference MAD as shown in Eq.(5.5) and Eq.(5.6) below and then apply the searching methods such as the 2D-logarithmic search, the three step search or the modified conjugate direction search that are described briefly in the following.

$$MSE(l,j) = \frac{1}{MN} \sum_{m=1}^{M} \sum_{n=1}^{N} [i(m,n) - i_{k-1}(m+l,n+j)]^2 \tag{5.5}$$

$$MAD(l,j) = \frac{1}{MN} \sum_{m=1}^{M} \sum_{n=1}^{N} |i(m,n) - i_{k-1}(m+l,n+j)| \tag{5.6}$$

where $-dm \leq l,j \leq +dm$

The 2D-logarithmic search is based on evaluating the mean square error and tracking the direction of minimum distortion. In each step, five points (North, East, South, West and Center) are checked. The minimum point will be chosen as a new Center. The distance between the search points is reduced if the minimum is Center itself or at the boundary of the search area.

The three step search is closely related to the 2D-logarithmic search except for the nine tested points (N, NE, E, SE, S, SW, W, NW and C) and based on the mean absolute difference. The first step is taken to surround the original point. In a second step, searching points are spaced less coarsely around the first approximation (the first new Center) and repeated until new Center is achieved by itself. Then, the third step gives the final displacement vector with least searching space.

The conjugate direction search procedure (see Fig.5.3) tracks the direction of minimum distortion D(i,j) which is defined by the mean absolute difference MAD(i,j) criterion. In a first step, the minimum in the x direction is determined by computing D(i-1,j), D(i,j) and D(i+1,j). Suppose D(i+1,j) is the smallest, calculate D(i,j), D(i+1,j) and D(i+2,j) next step. Continue searching until the

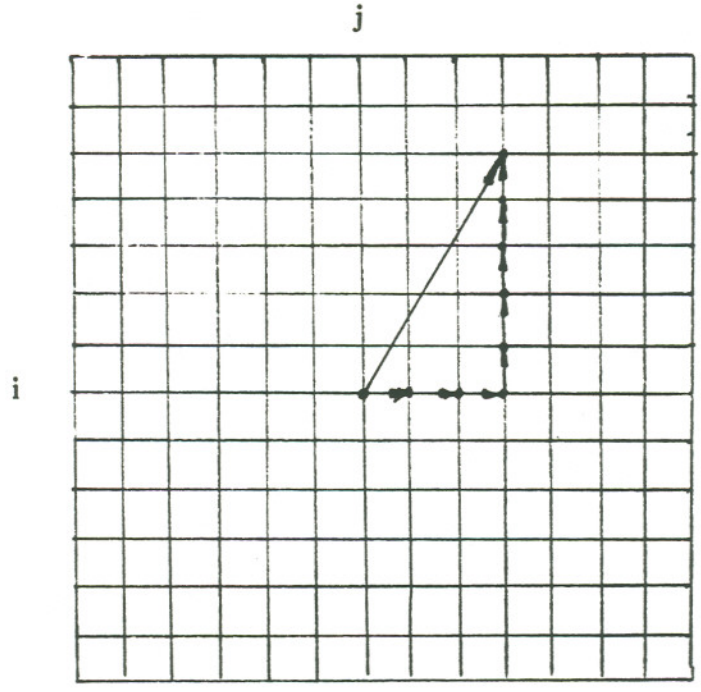


Fig.5.3 Illustration of conjugate direction search procedure

smallest value is positioned between two higher values. Then start searching by the same procedure but in y direction. The distance from point (i,j) and minimum point ($i+i_{min}$, $j+j_{min}$) represents the magnitude of the displacement vector and (i_{min} , j_{min}) is the direction of the vector.

The accuracy of the described block matching algorithms, is recognized, depends on the size of the block. The smaller the size, the higher the accuracy, the more the computation. The resolution of block matching algorithms is limited to 0.5 picture element. To reach the higher accuracy, correlation technique[57] is required to describe the frame difference such as

$$CF = \frac{R_{i_k\ ,\ i_{k-1}}(D)}{\sqrt{R_{i_k\ ,\ i_k}(0) R_{i_{k-1}\ ,\ i_{k-1}}(0)}} \tag{5.7}$$

where $R_{i_k,i_{k-1}}(D)$ is cross correlation with the dispacement D between kth frame and k-1th frame; $R_{i_k,i_k}(0)$ and $R_{i_{k-1},i_{k-1}}(0)$ are auto correlations without displacements in frame k and frame k-1, respectively.

5.1.2. Recursive displacement estimation algorithms

To improve the estimation accuracy, recursive estimation algorithm is often used although it requires more computations than the block matching algorithm. In this method, a new improved estimate $\hat{D}_{n+1}$ is produced by an initial estimate $\hat{D}_n$ and update term of iteration u_j according to

$$\hat{D}_{j+1} = \hat{D}_j + u_j \tag{5.8}$$

The converged value of $\hat{D}_j$ provides displacement. The iterations can be executed either in horizontal direction (from pixel to pixel) or vertical direction (from line to line) or temporal direction (from frame to frame). The same as before, a function of the displaced frame difference DFD can be used as a criterion for calculating the estimate $\hat{D}_{j+1}$

$$DFD(x,y,\hat{D}_j) = i_k(x,y) - i_{k-1}(x-d\hat{x}_j, y-d\hat{y}_j) \tag{5.9}$$

If $d\hat{x}_j$, $d\hat{y}_j$ are not integrals, neighboring interpolation is required to evaluate $i_{k-1}(x-d\hat{x}_j, y-d\hat{y}_j)$ and DFD. According to linear two dimensional interpolation (Fig.5.4), the i_{k-1} will be

$$i_{k-1}(x-d\hat{x}_j, y-d\hat{y}_j) = (1-\Delta y)[(1-\Delta y)i_A + \Delta x i_B] + \Delta y[(1-\Delta x)i_C + \Delta x i_D] \tag{5.10}$$

Other simplified versions for the interpolation have also been proposed in order to reduce the implementation complexity.

For update terms, several searching algorithms have been published (Eq.(5.11)). To simplify the description of these algorithms only x component of the displacement vector is considered.

Netravali and Robbins[59] ($\epsilon = \frac{1}{1024}$)

$$u_j = -\epsilon \sum_M DFD(x,y,\hat{D}_j) \frac{\partial}{\partial x} i_{k-1}(x-d\hat{x}_j, y-d\hat{y}_j) \tag{5.11a}$$

Newton–Raphson[60]

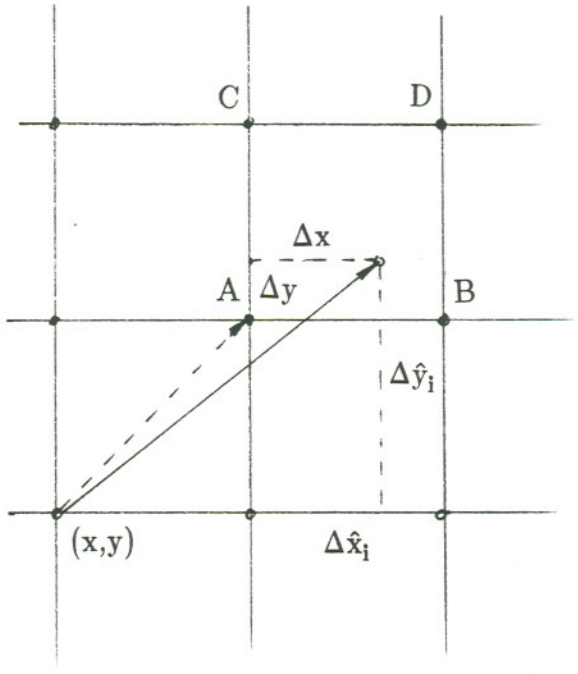


Fig.5.4 Interpolation for estimate frame difference

$$u_j = -\frac{\sum_M \frac{\partial}{\partial x}[DFD(x,y,\hat{D}_j)]^2}{\sum_M \frac{\partial^2}{\partial x^2}[DFD(x,y,\hat{D}_j)]^2} \tag{5.11b}$$

Cafferio and Rocca[61] ($\eta^2 = 100$)

$$u_j = -\frac{\frac{1}{2}\sum_M \frac{\partial}{\partial x}[DFD(x,y,\hat{D}_j)]^2}{\sum_M \left[\frac{\partial}{\partial x} i_{k-1}(x-\hat{dx},y-\hat{dy})\right]^2 + \eta^2} \tag{5.11c}$$

Bergmann[62]

$$u_j = -\frac{\frac{1}{2}\sum_M \frac{\partial}{\partial x}[DFD(x,y,\hat{D}_j)]^2}{\frac{1}{2}\sum_M \left[\frac{\partial}{\partial x} i_{k-1}(x-dx,y-dy)+\frac{\partial}{\partial x} i_k(x,y)\right]\frac{\partial}{\partial x} i_k(x,y)} \tag{5.11d}$$

Recursive algorithms allow a more accurate estimate and the cost is more complex computations. It should be also mentioned that a quick convergence might yield a noisy estimate. To achieve more accuracy, more time is needed. So, the choice of quick convergence and real-time computation requires a compromise.

5.1.3. Phase correlation estimation algorithm

The phase correlation method takes the advantage of the fact that the most of the information about the relative displacement vector between two images is

contained in the phase of their cross-power spectrum. The location of peak in correlation surface gives the displacement vector (see Fig.5.5). The computation starts from obtaining the two-dimensional Fast Fourier Transform (FFT) I_1 and I_2 from two images i_1 and i_2 with the same dimension N×M. Then, forming the cross-power spectrum to derive the phase difference matrix

$$e^{j\phi} = e^{j(\phi_1-\phi_2)} = \frac{I_1 I_2^*}{|I_1 I_2^*|} \tag{5.12}$$

And then, filtering by a weighting function H(f) and taking the inverse FFT of weighted phase matrix. The phase correlation function d is obtained by

$$d = \mathcal{F}^{-1}[He^{j\phi}] \tag{5.13}$$

[If the weighting function is in the form $|I_1 I_2^*|^{\alpha}$, the result will be

phase correlation

$$d = \mathcal{F}^{-1}[e^{j\phi}] \quad \text{when } \alpha = 0$$

or cyclic cross correlation

$$d = i_1 * i_2 = \mathcal{F}^{-1}[I_1 I_2^*] = d * \mathcal{F}^{-1}[|I_1 I_2^*|] \quad \text{when } \alpha = 1$$

Varying the parameter α between 0 and 1, the algorithm will be optimal for from a wide-bandwidth signal degraded by narrow-bandwidth noise to a narrow-bandwidth signal degraded by wide-bandwidth noise].

One of improved phase correlation estimation is block matching phase correlation. The procedure of this technique is to shift the entire source image by

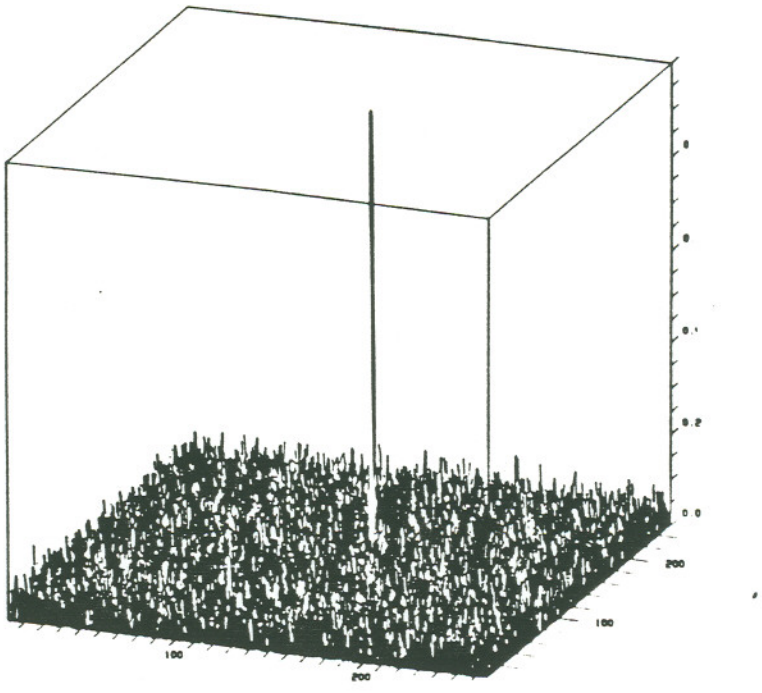


Fig.5.5 Illustration of phase correlation estimation scheme

each of the trial vectors in turn, which are determined by the phase correlation technique[63], then calculate the luminance difference between successive pictures with respect to pixels. The vector with the least error is assigned to that area. On the correlation surface, peaks correspond to the relative displacement of the images. With the aid of interpolating location of several prominent peaks, trial vectors are obtained with subpixel accuracy.

Obviously, this technique can give a wide shift range measurement. The accuracy was also found to be about 0.1 pixels[64]. However, it will be accompanied with a large computational need.

5.2. Algorithms for Estimation Based on Feature and Optic Flow

Based on human vision system, Gedanken experiment indicates that stereovision can be applied by superposition of multi monovision. For example, if we cover right eye and store the image recorded by left eye, then cover left eye and store the image recorded by right eye, these two monocular images we got are equivalent to one stereo image (Fig.5.6(a)(b)(c)). The same information will be given (Fig.5.6(d)) when we cover one eye and move our head sideways by the distance between two eyes. It is also fortunate to have the same effect (Fig.5.6(e)(f)) when we keep our head stationary and let the object move a certain distance or under the case of enlarging or shrinking the "space point" with the same scale distance jumping and moving. This gives the idea of evaluating at least two monocular frames as in stereo vision. To recover a moving stereo vision, not only two frames but multi frames with recording more than one aspect of an object

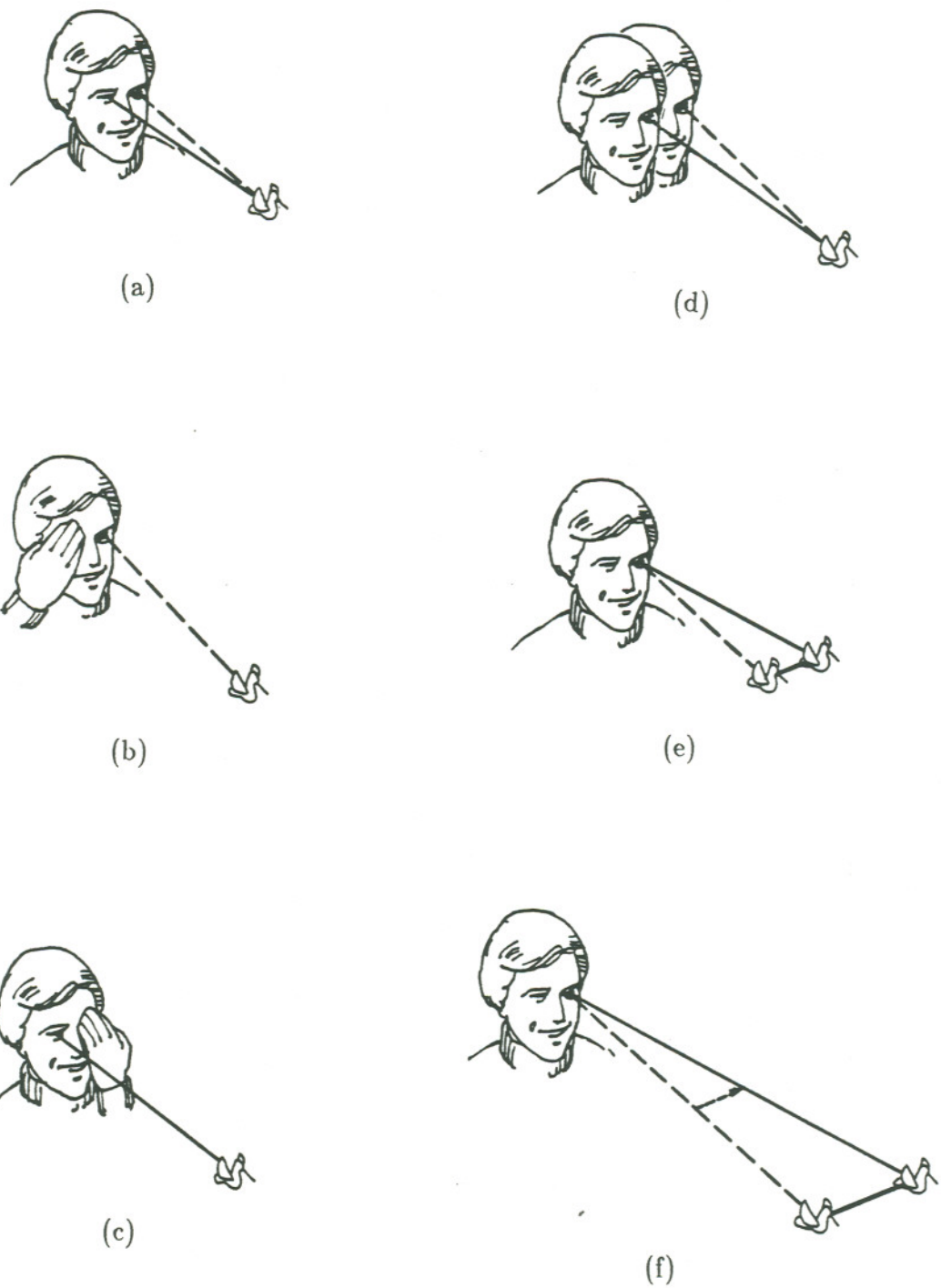


Fig.5.6 Vision system

are required. Another phenomenon is no matter whether the camera is in motion or the imaged objects are moving, the relative position and motion is the same in imaged scene between the camera and the object. So, when we discuss the motion, we can assume the camera is still while the target is moving.

Physically speaking, the relative motion may be characterized by a set of features or brightness patterns in the image. The procedure of perceiving the motion of the objects involves the analysis of the motion of features or brightness patterns. Feature based approach and optic flow based approach are two distinct branches in motion estimation nowadays. Feature based motion estimation is based on extracting two dimensional features in the images which correspond to three dimensional object features in the scene, such as points, lines or curves from corners or boundaries of surfaces in an image. Optic flow based motion estimation is based on computing the optic flow (the distribution of apparent velocities of movement of brightness patterns) in an image.

5.2.1. Feature based motion estimation

Two frames and multiple of special points such as corners can be used to determine the motion. Suppose moving object can be considered as a rigid body, the motion can then be described as translation or rotation. Assume (x,y,z) and (u,v,w) are the camera coordinate system and the object coordinate system (Fig.5.7)[65] respectively. $\underline{F}$ is the vector of a prominent point on the moving body, based on object coordinate system. The vector $\underline{C}$ of this prominent point with respect to the camera coordinate system may be derived by a

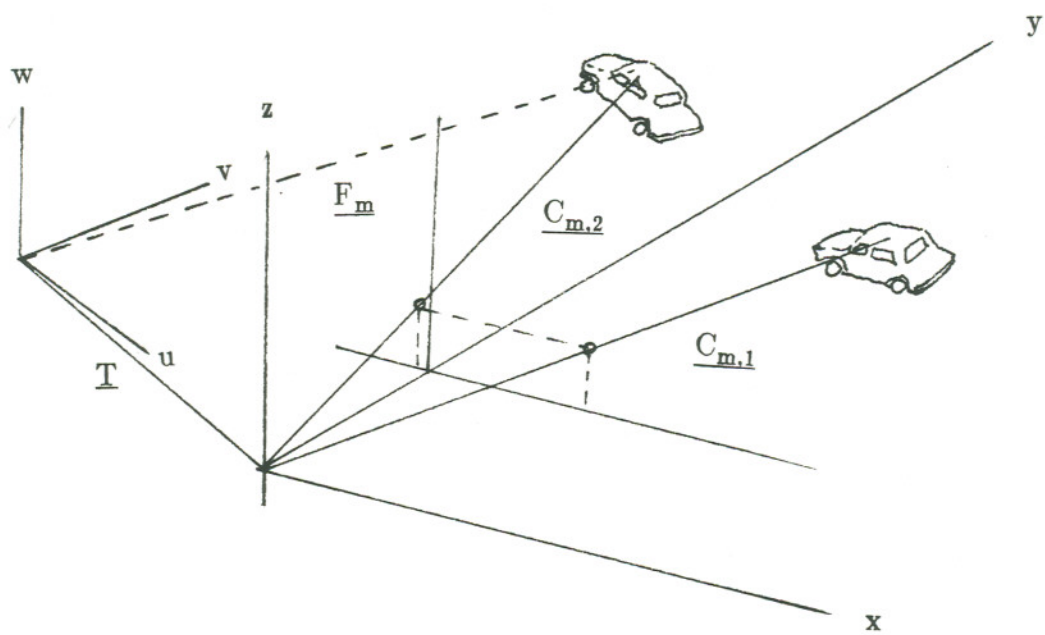


Fig.5.7 Camera coordinate system for deriving 3-D point configurations from a series of 2-D measurments

transformation

$$\underline{C}_{m,n} = (\underline{F}_m + \underline{T}_n)\underline{D}_n \tag{5.14}$$

where, m is point number; $\underline{T}$ is the translation vector which may be represented by the displacement between the origins of two frames; D is the rotation matrix required to align their axes. Since the object coordinate system moves together with the object, $\underline{F}$ and $\underline{D}$ will be a function of time, that is denoted by n.

Let the two coordinate system coincide in frame 1

$$\underline{C}_{m,1} = \underline{F}_{m,1} \tag{5.15}$$

The $\underline{C}$ in frame 2 will be represented by

$$\underline{C}_{m,2} = (\underline{F}_{m,1} + \underline{T})\underline{D} = (\underline{C}_{m,1} + \underline{T})\underline{D} \tag{5.16}$$

Now, what we should do is to find corresponding points, solve the equations and to get the solution of $\underline{T}$ and $\underline{D}$. Roach and Aggarwal[66] showed that five points in two views are required to recover these parameters and determine the structure. Nagel[65] reduced the dimension of the search space through the elimination of unknown variables, i.e. eliminating the translation vector and solving the rotation matrix separately. 3-D lines are also used as features in the estimation of structure and motion. But main task is still to get the solution of translation vector and rotation matrix.

As shown above, establishing and maintaining the correspondence available between features from one image and from the next image in a sequence is one

requirement of feature based approaches. The task to drop away hidden features and false features and to set up feature correspondence is difficult. This area is still in its infancy and needs to be developed. Sensitivity to noise is also a problem although it can be decreased by using more than the required minimum number of features in an iterative least squares technique.

5.2.2. Optic flow based motion estimation

The optic flow techniques[67] rely on the local spatial and temporal derivatives. This approach can be decomposed into two steps: 1) estimate optic flow field (image plane velocities) by calculating image intensity changes, and 2) use optic flow map to compute 3-D motion and structure. Since the intensity can be assumed not to change along the path of motion, the motion constraint comes out

$$i(x+\Delta x,y+\Delta y,t+\Delta t) = i(x,y,t) \tag{5.17}$$

where Δx, Δy and Δt are small.

Applying first order Taylor approximation and taking the limit as $\Delta t \to 0$, Eq.(5.17) can be approached to

$$\frac{\partial i}{\partial x}\frac{dx}{dt} + \frac{\partial i}{\partial y}\frac{dy}{dt} + \frac{\partial i}{\partial t} = 0$$

Denoted as

$$i_x u + i_y v + i_t = 0 \tag{5.18}$$

u and v describe the instantaneous displacement velocity of image motion. Knowing u and v, the motion is determined. Since there are two unknowns in Eq.(5.18), it is not sufficient to specify the optic flow uniquely. Therefore, additional assumption is required, such as constant optic flow over an entire segment of the image or smooth one with similar velocities in neighboring points etc.

Nagel[68] assumed the displacement is constant for entire intensity structure. He added additional equation by estimating the minimum of the squared differences between the intensity structures observed at different time.

$$\iint [i(x,y,t+\Delta t) - i(x-u\Delta t, y-v\Delta t, t)]^2 \, dxdy = \text{minimum} \tag{5.19}$$

He also noticed the characteristics of a "gray value corner" as the location of maximum planar curvature can be formulated by $i_{xy}=0$. Finally, he used recursive method to acquire the velocity u and v.

Haralick and Lee[69] interpreted the motion constraint equation as the intersection line of the iso-intensity contour plane at time t_0 and t_1. To match these points along the line, the intensity match and the first intensity derivative match are required and higher order partial derivatives can be neglected. Then the additional equation Eq.(5.20) becomes Eq.(5.21) by the Taylor expansion

$$i_x(x_0,y_0,t+\Delta t) = i_x(x_0-u\Delta t, y_0-v\Delta t, t) \tag{5.20a}$$

$$i_y(x_0,y_0,t+\Delta t) = i_y(x_0-u\Delta t, y_0-v\Delta t, t) \tag{5.20b}$$

$$i_t(x_0,y_0,t+\Delta t) = i_t(x_0-u\Delta t,y_0-v\Delta t,t) \tag{5.20c}$$

$$i_{xx}u + i_{xy}v + i_{xt} = 0 \tag{5.21a}$$

$$i_{yx}u + i_{yy}v + i_{yt} = 0 \tag{5.21b}$$

$$i_{tx}u + i_{ty}v + i_{tt} = 0 \tag{5.21c}$$

This overdetermined system can be solved by the pseudo-inverse formalism as follows

$$\begin{bmatrix} i_x & i_y \\ i_{xx} & i_{xy} \\ i_{yx} & i_{yy} \\ i_{tx} & i_{ty} \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = - \begin{bmatrix} i_t \\ i_{xt} \\ i_{yt} \\ i_{tt} \end{bmatrix} \tag{5.22}$$

short for

$$\mathbf{A}\begin{bmatrix} u \\ v \end{bmatrix} = -\mathbf{B}$$

$$\begin{bmatrix} u \\ v \end{bmatrix} = -[\mathbf{A}^T\mathbf{A}]^{-1}\mathbf{A}^T\mathbf{B} \tag{5.23}$$

Meanwhile, Tretiak and Pasten[70] differentiated the motion constraint equation with respect to x and y under constant velocity assumption. They also used the pseudo-inverse formalism to get the solution of the overdetermined equations.

Practically, the velocity might not be a constant. Linear optic flow assumption is generally used in that case. This relaxation approach is equivalent to

neglecting higher order (>2) terms when Taylor expansion is employed:

$$i_x(x_0+\Delta x,y_0+\Delta y,t)u(x_0+\Delta x,y_0+\Delta y) + i_y(x_0+\Delta x,y_0+\Delta y,t)v(x_0+\Delta x,y_0+\Delta y)$$

$$+ i_t(x_0+\Delta x,y_0+\Delta y,t) = 0 \tag{5.24}$$

We get

$$i_x u + i_y v + i_t = 0 \tag{5.24a}$$

$$i_{xx}u + i_{yx}v + i_x u_x + i_y v_x + i_{tx} = 0 \tag{5.24b}$$

$$i_{xy}u + i_{yy}v + i_x u_y + i_y v_y + i_{ty} = 0 \tag{5.24c}$$

$$i_{xx}u_x + i_{yx}v_x = 0 \tag{5.24d}$$

$$i_{xy}u_x + i_{xx}u_y + i_{yy}v_x + i_{yx}v_y = 0 \tag{5.24e}$$

$$i_{xy}u_y + i_{yy}v_y = 0 \tag{5.24f}$$

Since the coefficient matrix of linear equations does not have a full rank of six, it requires additional information to determine all unknowns. Nagel[67] imposed specific gray value corner conditions and developed the Taylor series up to second order terms. He supposed $v_x - u_y = 0$, a requirement weaker than constant velocity, but still results constant velocity as a solution.

Horn and Schunk[71] assumed motion field varies smoothly in most image area. They introduced this smoothness condition by minimizing the error in optic flow as expressed below.

$$\iint [(i_y u + i_y v + i_t)^2 + \alpha^2(u_x^2 + u_y^2 + v_x^2 + v_y^2)]\, dxdy = \text{minimum} \quad (5.25)$$

By means of calculus of variation, after differentiating with respect to u and v and letting them zeros, one will get iterative answer (Eq.(5.27)) by defining Eq.(5.26).

$$u_x^2 + u_y^2 = (u - u_{ave})^2 \quad (5.26a)$$

$$v_x^2 + v_y^2 = (v - v_{ave})^2 \quad (5.26b)$$

$$u = u_{ave} - i_x \frac{i_x u_{ave} + i_y v_{ave} + i_t}{\alpha^2 + i_x^2 + i_y^2} \quad (5.27a)$$

$$v = v_{ave} - i_y \frac{i_x u_{ave} + i_y v_{ave} + i_t}{\alpha^2 + i_x^2 + i_y^2} \quad (5.27b)$$

In addition, Yachida[72] extended Horn and Schunck's smoothness constraint by considering both spatial and temporal neighborhood.

Nagel[73] took into account occluding edges and proposed a modified Horn and Schunck's smoothness criterion by introducing a weight matrix. Enkelmann[74], Clazer[75] and Anandan[76] used Guassian low pass pyramid, Guassian pyramid and Laplacian pyramid bandpass respectively to provide processing from a coarse level to fine level. Mitiche, Wang and Aggarwal[77] computed optic flow by using multi- constraint methods, and so on. The above approaches deal with images with a single moving objects. The determination of multi moving objects has also been attempted.

Since the optic flow approaches require the evaluation of partial derivatives of image gray values, the presence of noise becomes a problem. The higher the order of partial derivatives, the more is the noise sensitivity. If the optic flow changes sharply or is discontinuous, the violations of the continuity assumption will introduce poor estimation.

5.3. Improved motion estimation

Each of the methods mentioned above has its own advantages and disadvantages. For the requirement of real time processing and accurate estimation, an adaptive and compromised method has to be recommended. The improved motion estimation below is one of reasonable selections.

The improved motion estimation system consists of both motion detection and motion estimation. The motion detection includes frame difference, threshold and 3-D lowpass filter. At first, frame difference and threshold reduce the noise and pick the slow motion area out. 3-D lowpass filter successively provides information of medium motion or fast motion. Sequentially, under different conditions, different motion estimation methods are applied to get their full advantages. The motion estimation includes optic flow based estimation for medium motion area and correlation search for fast motion area where block matching is used to simplify the calculation. Fig.5.8 gives the block diagram of the estimation system.

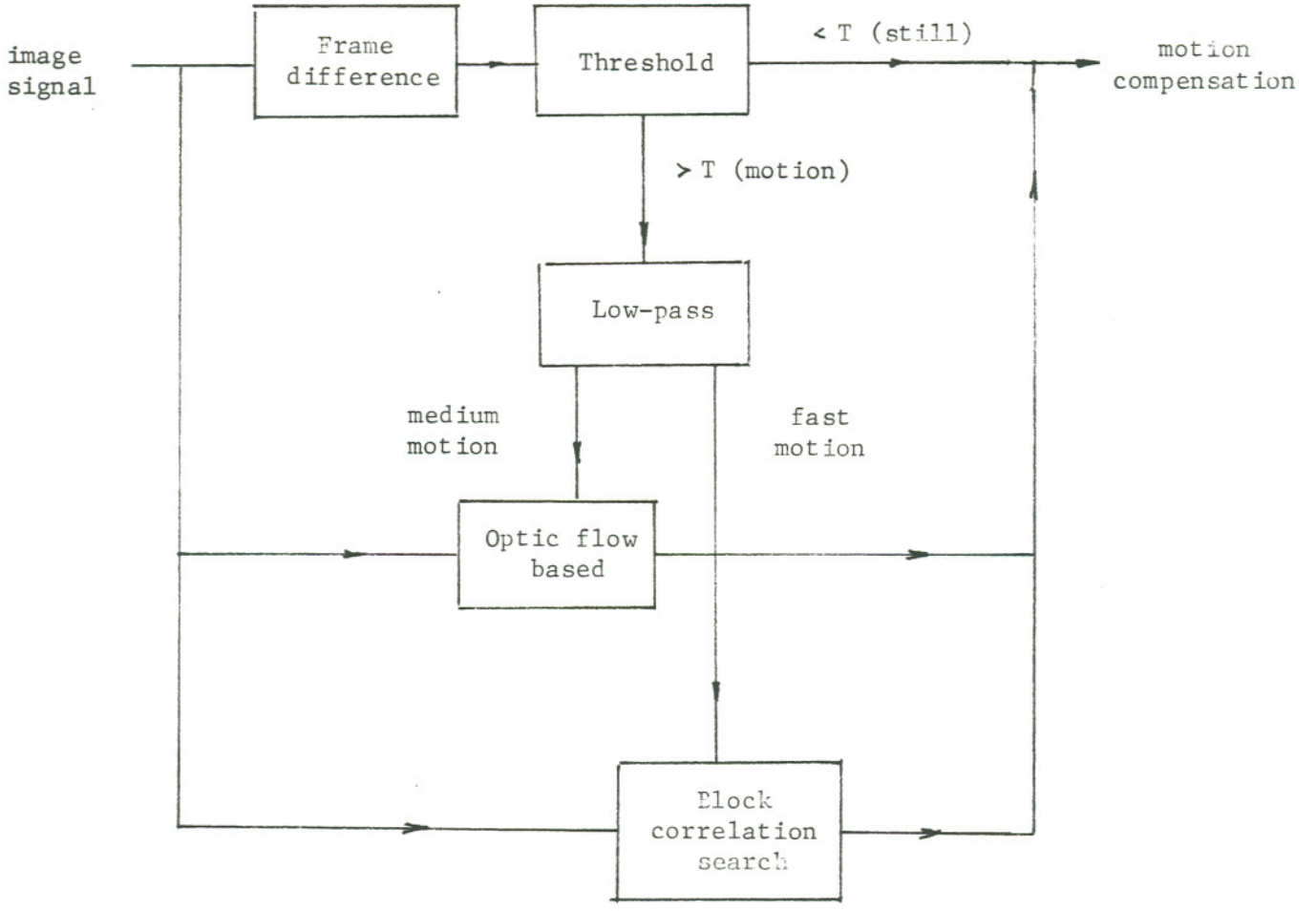


Fig.5.8 Block diagram of improved estimation system

5.3.1. Motion Detection

In motion detection, a series of filtering has been adopted. First, slow motion is picked up from pictures by thresholding successive frame difference on a pixel. Zero value indicates the region is still area. Non zero portion explicates there are medium or fast movements in that area. Second, in the region with non zero difference, a 3-D low pass filter is applied to the difference signal. Therefore, the medium motion portion is filtered out, and the fast portion can be obtained by subtracting slow and medium portion from the entire area.

To simplify 3-D low pass filter implementation, a separable filter is adopted, which includes a 9-point spatial filter and a group of 7-point temporal filters. One of practical spatial filters is the symmetrical filter. Its frequency response $H_s(f_x,f_y)$ corresponds to $h_s(m\Delta x,n\Delta y)$

$$h_s^{MN}(m\Delta x,n\Delta y) = \begin{cases} h_0 & (m,n)=(M,N) \\ h_1 & (m,n)=(M\pm1,N),(M,N\pm1) \\ h_2 & (m,n)=(M\pm1,N\pm1),(M\mp1,N\pm1) \\ 0 & \text{otherwise} \end{cases} \tag{5.28a}$$

is given by

$$\begin{aligned} H_s^{MN}(f_x,f_y) &= \sum_m \sum_n h_s(m\Delta x,n\Delta y)\, e^{-j2\pi(xf_x+yf_y)} \\ &= [h_0+2h_1\cos(2\pi\Delta xf_x)+2h_1\cos(2\pi\Delta yf_y)+4h_2\cos(2\pi\Delta xf_x)\cos(2\pi\Delta yf_y)] \\ &\quad \times e^{-j2\pi(M\Delta xf_x+N\Delta yf_y)} \end{aligned} \tag{5.28b}$$

where m, n, M and N are integers; Δx and Δy are spaces between two nearest points in horizontal and vertical direction respectively; $(M\Delta x, N\Delta y)$ is the point of interest. A typical symmetrical filter is an averaging filter. When $h_0=h_1=h_2=\frac{1}{9}$, its frequency response (Fig.5.9(a)) is

$$H_s^{MN}(f_x,f_y) = \sum_m \sum_n h_s(m\Delta x, n\Delta y)\, e^{-j2\pi(xf_x+yf_y)}$$

$$= \frac{1}{9}[1+2\cos(2\pi\Delta x f_x)+2\cos(2\pi\Delta y f_y)+4\cos(2\pi\Delta x f_x)\cos(2\pi\Delta y f_y)]$$

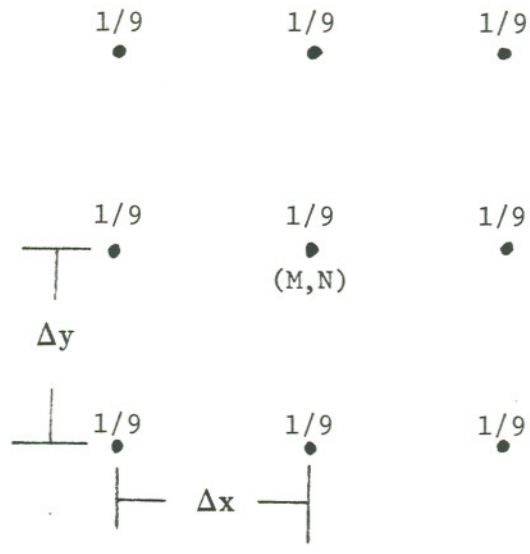
$$\times e^{-j2\pi(M\Delta x f_x+N\Delta y f_y)} \qquad (5.29)$$

If we choose $h_0=\frac{1}{2}$, $h_1=\frac{1}{12}$ and $h_2=\frac{1}{24}$, the picture itself will be sharper than that of averaging filter.

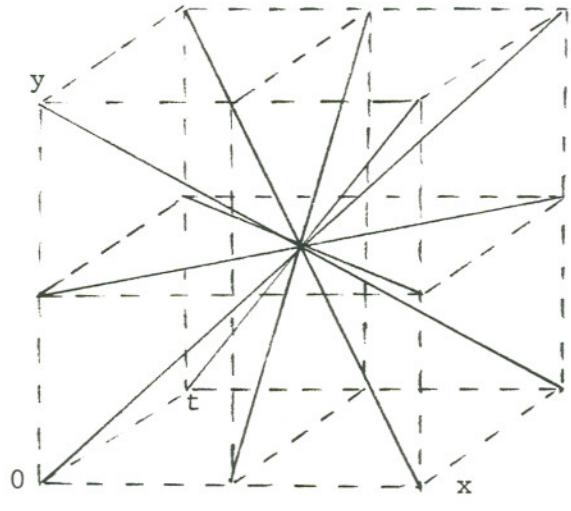
Temporal filter group can be considered as a 1-D filter applied on different directions. An example of eight-filter group is shown in Fig.5.9(b). These directions are also useful for motion estimation. A 1-D low pass filter can be designed by existing filter design programs. Its frequency response denotes $H_t^{xy}(e^{j2\pi f_t})$. The frequency response of the filter group is expressed as

$$H_{tg}(f_x,f_y,f_t) = H(e^{j2\pi(uf_x + vf_y + f_t)}) \qquad (5.30)$$

where u and v are velocities x/t and y/t in horizontal and vertical direction respectively, and they determine the temporal orientation of the filters.



(a) 9-point average spatial filter



(b) Location of a group of temporal filters

Fig.5.9 3-D low pass filter in impoved estimation system

Since the separate filters are used in cascade, the frequency response of the 3-D low pass filter is obtained by multiplication of these two separate filter frequency responses.

$$H(f_x,f_y,f_t) = H_s(f_x,f_y)\, H_{tg}(f_x,f_y,f_t) \tag{5.31}$$

5.3.2. Motion Estimation

After motion detection, motion estimation is followed. In the slow motion area, considering a stationary picture, motion estimation is not necessary. In the medium motion area, since the movement is not so quick, small changes both in spatial and temporal directions approximated by the derivative and the gradient are available and effective. Therefore, the optic flow based motion estimation is adopted. Using the polynomial intensity assumption[81], motion structure and vector are determined. In fast motion area, we search for translation vector first, then the rotation angle and axis as described below.

For translation vector, searching starts from the initial correlation point and is done between two frames. The initial point is selected from two reference points by the minimum mean absolute correlation difference. The two reference points in successive images are located on the motion direction indicated by the motion detection and on the margin between the medium motion area and the fast motion area. After that, find the minimum mean absolute correlation difference from those corresponding with the point of interest (O) and the initial point (C), point (O) and initial left closest point (W), point (O) and initial right

closest point (E), point (O) and initial up closest point (N) and point (O) and initial down closest point (S), respectively. Then, create a new initial point (C_1) with the minimum difference. In the next step, only 4 differences need to be considered. Follow the same procedure as the previous step until the new initial point (C_{i+1}) is the previous initial point (C_i) itself. Therefore, $\underline{OC_i}$ is approximated as the motion vector for that block. Let us explain the searching procedure by the example shown in Fig.5.10. The point O in kth frame is the point of interest with gray value $i(m,n,t_k)$. The other points are the reference points located on the previous frame. In the first step, select two reference points. Suppose the direction indicated by the motion detection is horizontal-temporal direction, the two reference points are C_+ and C_-. which are l apart from O along x axis with the gray value $i(m+l,n,t_{k-1})$ and $i(m-l,n,t_{k-1})$. Calculate and compare with the mean absolute correlation differences between these two points with respect to point O. Choose the point with the minimum mean absolute correlation difference as the new initial point (assume here the point is C_+). In the second step, calculate MACDs on the four points (E, S, W and N) around the initial point O and compare MACDs on five points (C, E, S, W and N). The point with minimum MACD is selected to be the new initial point (assume here the point is N and is denoted as C_1). In the third step, MACDs only on three points (E_1, W_1 and N_1) need to be calculated and four points (E_1, W_1, N_1 and C_1) need to be compared to get the new initial point. Follow this step until the new initial point is the initial point itself, the vector from the point of interest to the final initial point is the translation vector (represented by $\underline{OC_{10}}$ in the example).

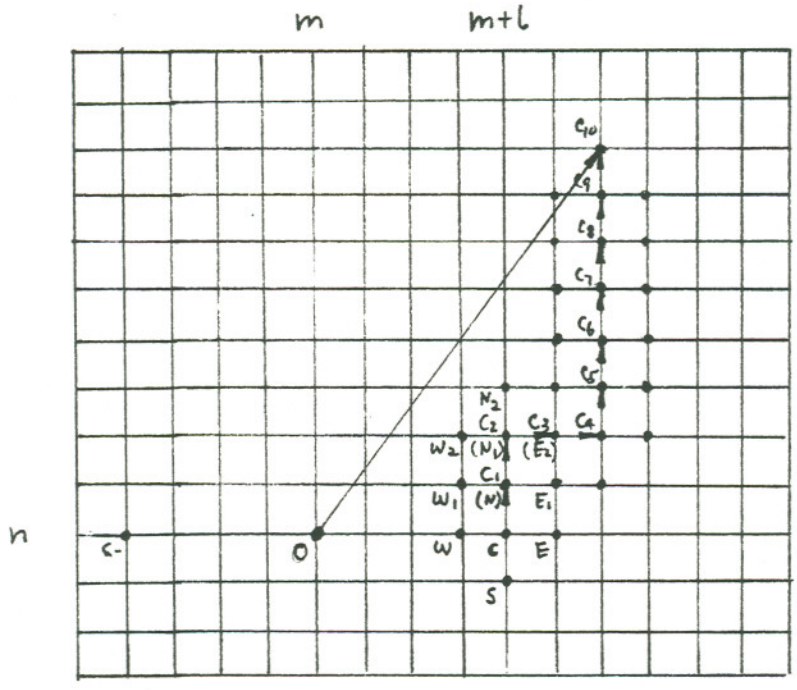


Fig.5.10 Searching procedure in improved estimation system

To simulate rotation of an object usually needs new information of the object such as its 3-D shape. So, to describe rotation roughly, slow rotation is assumed. It is feasible to trace the rotation by considering the correlation between compressed block b_1 ($b_{1x} \times b_{1y}$) in the frame of interest and the same size block b_0 in the previous frame. The processing is to compress block b of size $b_x \times b_y$ to b_1 of size $b_{1x} \times b_{1y}$ symmetrically as shown in Fig.5.11. According to rotation axes (a) and rotation angle (α) (two rotation axes centered x and centered y and four rotation angle $\pm 30°$ and $\pm 15°$ are adopted here). Sequentially, testing the different rotation models in turn with different rotation axes and angles; getting intensity with corresponding linear weight and checking the mean absolute correlation difference between the created blocks b_1s and previous block b_0, the rotation axis and rotation angle with the minimum MACD is assigned to this block. Since the object might be far from the camera and the image, the gray value would have no change or be slightly varied so that constant or small change weight is suggested. For instance, assume rotation angle is 30° with respect to axis y and the weight is constant. First, compute mean absolute correlation difference in the 87% center region of entire area in the previous frame k-1. Next, compress the entire area in frame k into small area with the rate of 0.87. Then, linearly weight (constant now) the intensity proportional to the coefficient shown in Fig.5.11 along the axis perpendicular to the rotation axis, calculate the mean absolute correlation difference based on these new gray values in the compressed region, and change the angle or axis until all MACD done. Finally, compare these MACDs in frame k with MACD in frame k-1. The optimal match model is the rotation estimation model.

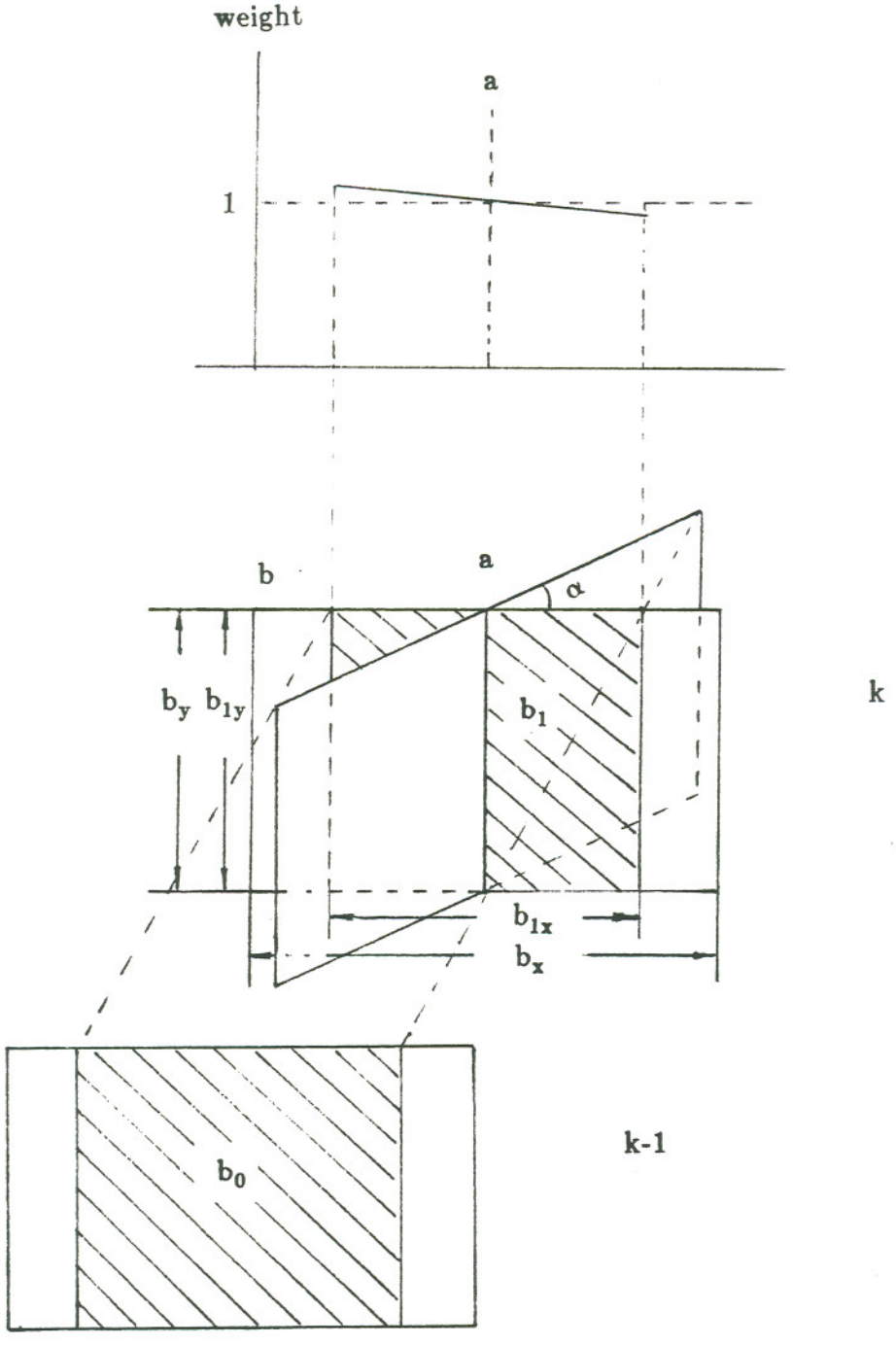


Fig.5.11 Rotation scheme in improved estimation system

In summary, because of solving the problem under different conditions, this method can get more accurate and reasonable motion estimation with less computational load. Since it requires series and parallel architectures to meet the needs of real time processing, hardware implementation will be more complicated.

CONCLUSION

High definition television is expected as next generation medium because it provides a high quality picture and sound. Since the signal bandwidth of the compatible HDTV system is restricted by the current TV standard, advanced signal processing techniques are required to achieve picture improvements. A great deal of the activity devoted to the development of HDTV is on motion estimation and motion compensation. The aspects of the motion estimation system are receiving a considerable amount of study and were discussed here.

A new appropriate motion estimation system for HDTV was proposed in the thesis. It consisted of simple detection and estimation. Since the processing was carried out under different conditions, different estimation methods were applied to their full advantages. Block processing, correlation search and optic flow method made the motion estimation more accurate without complex computation. The method is applicable but must await further testing.

The efforts that are underway include further development of the motion estimation, multi-target motion estimation and effective implementation.

REFERENCES

[1] R. K. Jurgen, "Chasing Japan in the HDTV Race." IEEE Spectrum, Oct. 1989

[2] M. J. Mackall, "Likely Costs of Consumer Advanced Television (ATV) Technology," IEEE Trans. Consumer Electronics, Vol.35, No.2, May 1989

[3] J. O. Drewery, "The Zone Plate as a Television Test Pattern," SMPTE J., Nov. 1979

[4] T. Fukinuki & Y. Hirano, "The To-and-Fro Zone Plate (TFZP) Method for Observing Frequency Characteristics in Three Dimensions," SMPTE J., Sept. 1986

[5] P. Saraga, "Compatible High-Definition Television," Electronics & Communication Engineering J., Jan/Feb. 1989

[6] T. Reuter, "Motion Adaptive Downsampling of High Definition Television Signal," SPIE, Vol.594, Image Coding (1985)

[7] N. T. Graarger, "A Note on the Multidimensional Sampling Theorem," Proc. IEEE, Feb. 1972

[8] P. S. Carnt & G. B. Townsend, *Colour Television*, Ififfe Bks. Ltd., London, 1968

[9] E. Dubois & W. F. Schreiber, "Improvements to NTSC by Multidimensional filtering," SMPTE J., June 1988

[10] D. Teichner, "Three-Dimensional Pre- and Post- Filtering for PAL TV Signals," IEEE Trans. Consumer Electronics, Vol. CE-34, No.1, Feb. 1988

[11] H. J. Tonge, "Image Processing for High Definition Television," IEEE Trans. Circuits & Systems, Vol CAS-34, No.11, Nov. 1987

[12] Farondja & Roizen, "Improving NTSC to Achieve Near - RGB Performance," SMPTE J., Aug. 1987

[13] T. Fukinuki & Y. Hirano, "Extended Definition TV Fully Compatible with Existing Standards," IEEE Trans. Communication, Vol. COM-32, No.8, Aug. 1984

[14] T. Fukinuki, et.al, "Fully Compatible EDTV for Improving both Y and Color Signals by Using A Single New Subcarrier," IEEE Trans. consumer Electronics, Vol.34, No.3, Aug. 1988

[15] Y. Yasunoto, et.al, "An Extended Definition Television System Using Quadrature Modulation of the Video Carrier with Inverse Nyquist Filter," IEEE Trans. Consumer Electronics, Vol. CE-33, No.3, Aug. 1987

[16] C. Hentschel, "Linear and Nonlinear Procedures for Flicker Reduction," IEEE Trans. Consumer Electronics, Vol. CE-33, No.3, Aug. 1987

[17] R. Hopkins, "Advanced Television Systems," IEEE Trans. Consumer Electronics, Vol.34, No.1, Feb. 1988

[18] M. A. Isnardi, C. B. Dieterich & T. R. Smith, "Advanced Compatible Television: A Progress Report," SMPTE J., July 1989

[19] Y. Ninoniya, et.al, "A Single Channel HDTV Broadcast System --- The MUSE," NHK Laboratories Note 304, Tokyo, Japan, Sept. 1984

[20] Rossi, Goldberg & McMann, "A Compatible HDTV Broadcast System," NAB Convertion, Proc. 38th Annual Broadcast Eng. Conf., Las Vegas, April, 1984

[21] Lo Cicero, Pazarci & Rzeszewski, "A Compatible High-Definition Television System (SLSC) with Chrominance and Aspect Ratio Improvements," SMPTE J., May, 1985

[22] W. E. Glenn & K. G. Glenn, "HDTV Compatible Transmission System," 128th SMPTE Technical Conf., N.Y., Oct. 1986

[23] R. J. Iredale, "A Proposal for a New High-Definition NTSC Broadcast Protocol," SMPTE J., Oct. 1987

[24] M. J. J. Annegarn, et.al, "HD-MAC: A Step Forward in the Evolution of Television Technology," Philips Technicl Review, Vol.43, No.8, 1987

[25] M. Tsingberg, "ENTSC Two-Channel Compatible HDTV System," IEEE Trans. Consumer Electronics, Vol. CE-33, No.3, Aug. 1987

[26] Isnardi, et.al, "A Single Channel, NTSC Compatible Widescreen EDTV System," 3rd Int. Colloquium on Advanced Television Systems, Ottawa, Oct. 1987

[27] J. L. E. Baldwin, "Enhancing Television --- An Evolving Scene," SMPTE J., May 1988

[28] G. Schamel, "Pre- and Post Filtering of HDTV Signals for Sampling Rate Reduction and Display Up - Conversion," IEEE Trans. Circuits & Systems, Vol. CAS-34, No.11, Nov. 1987

[29] W. R. Schreiber, et.al, "Channel - Compatible 6-MHz HDTV Distribution Systems," SMPTE J., Jan. 1989

[30] P. Robert, M. Lamnabhi & J. J. Lhuillier, "Advanced High Definition 50 to 60 Hz Standards Conversion," 130th SMPTE Technical Conf., N.Y., Oct. 1988

[31] K. Sunada, et.al, "High Picture Quality TV Receiver with IDTV System," IEEE Trans. Consumer Electronics, Vol.34, No.4, Nov. 1988

[32] D. Nasse & J. Chatel, "Toward a World Studio Standard for High Definition Television," SMPTE J., June 1989

[33] W. E. Bretl, "3×NTSC—A 'Leapfrog' Production Standard for HDTV," SMPTE J., Mar. 1989

[34] R. J. Iredale, "HD—PRO™: A New Global High-Definition Video Production Format," SMPTE J., June 1989

[35] M. Tsingberg, F. Azadegem & E. Fisch, "Introduction of A NTSC Compatible HDTV System --- HDS/NA," IEEE Trans. Consumer Electronics, Vol.35, No.3, Aug. 1989

[36] W. Bretl, et.al, "Spectrum - Compatible High-Definition Television Transmission System," SMPTE J., Oct. 1989

[37] S. Sabri, D. Lemay & E. Dubois, "A Modular Digital Video Coding Architecture for Present and Advanced TV Systems," SMPTE J., July 1989

[38] Y. Yasumoto, et.al, "A Wide Aspect Ratio Television System with Full NTSC Compatibility," IEEE Trans. Consumer Electronics, Vol.34, No.1, Feb. 1988

[39] S. Kageyama, et.al, "An NTSC Compatible Wide Screen Television System with Evolutionary Extendibility," IEEE Trans. Consumer Electronics, Vol.34, No.3, Aug. 1988

[40] J. G. Raven, "High Definition MAC: The Compatible Route to HDTV," IEEE Trans. Consumer Electronics, Vol.34, No.1, Feb. 1988

[41] T. Kurita, et.al, "A Practical IDTV System Improving Picture Quality for Nonstandard TV Signals," IEEE Trans. Consumer Electronics, Vol.34, No.3, Aug. 1988

[42] K. H. powers, "The Historical Development and Standardization of the NTSC Color System," 131st SMPTE Technical Conf. & Equipment Exhibit, Los Angeles, CA., USA, Oct. 1989

[43] S. Takayama, et.al, "Experiment with Second Generation EDTV System," 131st SMPTE Technical Conf. & Equipment Exhibit, Los Angeles, CA., USA, Oct. 1989

[44] J. Alvarez & A. Cavallerano, "A 1050 Line Interlaced HDTV Format," 131st SMPTE Technical Conf. & Equipment Exhibit, Los Angeles, CA., USA, Oct. 1989

[45] C. P. Markhauser, "Some Now Adaptive Comb Filters for the Improvement of NTSC Video Images," 131st SMPTE Technical Conf. & Equipment Exhibit, Los Angeles, CA., USA, Oct. 1989

[46] S. M. Weiss, "Selecting Appropriate HDTV production Standards for North America: Making the Puzzle Pieces Fit," 131st SMPTE Technical Conf. & Equipment Exhibit, Los Angeles, CA., USA, Oct. 1989

[47] B. D. Dayton, "A Hierarchical Approach to HDTV Production Standards," 131st SMPTE Technical Conf. & Equipment Exhibit, Los Angeles, CA., USA, Oct. 1989

[48] T. Jarske, et.al, "Quincunx Coding for Picture Memories," 131st SMPTE Technical Conf. & Equipment Exhibit, Los Angeles, CA., USA, Oct. 1989

[49] S. Morikura, et.al, "PFWM Fiberoptic Transmission System for HDTV," 131st SMPTE Technical Conf. & Equipment Exhibit, Los Angeles, CA., USA, Oct. 1989

[50] K. R. Field, "Standards Conversion for HDTV - A Progress Report," 131st SMPTE Technical Conf. & Equipment Exhibit, Los Angeles, CA., USA,

Oct. 1989

[51] S. C. Hsu, "The Kell Factor: Past and Present," SMPTE J., Feb. 1986

[52] G. J. Tonge, "The Television Scanning Process," SMPTE J., July 1984

[53] D. E. Dudgeon & R. M. Mersereau, *Multidimensional Digital Signal Processing*, Prentice-Hall, Englewood Cliffs, N.J., 1984

[54] D. Teichner, "Adaptive Filter Techniques for Separation of Luminance and Chrominance in PAL TV Signals," IEEE Trans. Consumer Electronics, Vol. CE-32, No.3, Aug. 1986

[55] R.E. Crochiere & L. R. Rabiner, *Multirate Digital Signal Processing*, Prentice-Hall, Englewood Cliffs, N.J., 1983

[56] *Advanced Topics in Signal Processing*, J. S. Lim & A. V. Oppenheim, Eds., Prentice-Hall, Englewood Cliffs, N.J., 1988

[57] H. G. Musmann, P. Pirsch & H. J. Grallert, "Advances in Picture coding," Proc. IEEE, Vol.73, No.4, Apr. 1985

[58] J. K. Aggarwal & N. Nandhakumar, "On the computation of Motion from Sequences of Images --- A Review," Proc. IEEE, Vol.76, No.8, Aug. 1988

[59] A. N. Netravali & J. D. Robbins, "Motion Compensated Television Coding --- Part 1," Bell Syst. Tech. J., Vol.58, Mar. 1979

[60] H. C. Bergmann, "Displacement Estimation Based on the Correlation of Image Segments," IEEE Proc. Int. Conf. on Electronic Image Processing, York, England, July 1982

[61] C. Cafforio & F. Rocca, "The Differential Method for Image Motion Estimation," Image Sequence Processing and Dynamic Scene Analysis, T. S. Huang, Ed. Berlin, Germany, Sringer-Verlag, 1983

[62] H. C. Bergmann, "Ein Schnell Knovergierendes Displacement --- Schatzverfahren fur die Interpolation von Fernsehbild Sequenzen," Ph.D Dissertation, Tech. Univ. of Hannover, Hannover, Germany, Feb. 1984

[63] J. J. Pearson, et.al, "Video-Rate Image Correlation Processor," SPIE, Vol.119, Application of Digital Image Processing (10cc1977)

[64] G. A. Thomas, "HDTV Bandwidth Reduction by Adaptive Subsampling and Motion - Compensation DATV Techniques," SMPTE J., May 1987

[65] H. -H. Nagel, "Representation of Moving Rigid Objects Based on Visual Observations," Computer, Aug. 1981

[66] J. W. Roach & J. K. Aggarwal, "Computer Tracking of Objects Moving in Space," IEEE Trans. Pattern Anal. Machine Intell., Vol. PAMI-1, No.2, Apr. 1979

[67] H. -H. Nagel, "On the Estimation of Optical Flow: Relations between Different Approaches and Some New Results," Artificial Intelligence 33 (1987)

[68] H. -H. Nagel, "Displacement Vectors Derived from Second Order Intensity Variations in Image Sequences," Computer Vision, Graphics and Image Process, Vol.21, 1983

[69] R. M. Haralick & J. S. Lee, "The Facet Approach to Optic Flow," Proc. of Image Understanding Workshop, Arlington, VA., 1983

[70] O. Tretiak & L. Pastor, "Velocity Estimation from Image Sequences with Second Order Differential Operators," Proc. Int. Conf. Pattern Recognition, Montreál, Que, 1984

[71] B. K. P. Horn and B. G. Schunck, "Determining Optical Flow," Artificial Intelligence 17 (1981)

[72] M. Yachida, "Determine Velocity Map by 3-D Iterative Estimator," Proc. of Int. joint Conf. on Artificial Intelligence, 1981

[73] H. -H. Nagel, "Constraints for the Estimation of Displacement Vector Fields from Image Sequences," proc. of Int. Joint Conf. on Artificial Intelligence, 1983

[74] W. Enkelmann, "Investigations of Multigrid Algorithms from the Estimation of Optical Flow Fields in Image Sequences," Proc. IEEE Computer Society Workshop on Motion: Representation and Analysis, May 1986

[75] F. Glazer, "Hierarchical Motion Detection," Ph.D Dissertation, COINS Dept. Univ. of Massachusetts, Amherst, MA, Feb. 1987

[76] P. Anandan, "A Unified Perspective on Computational Techniques for the Measurement of Visual Motion," Proc. 1st Int. Conf. Computer Vision,

London, England, June 1987

[77] A. Mitiche, Y. F. Wang & J. K. Aggarwal, "Experiments in Computing Optical Flow with the Gradient - Based, Multiconstraint Method," Pattern Recognition, Vol.20, No.2, 1987

[78] E. C. Hildreth, "Computations Underlying the Measurement of Visual Motion," Artificial Intelligence 23 (1984)

[79] W. Richards, "Structure from Stereo and Motion," J. Opt. Soc. Amer. A, Vol.2, No.2, Feb 1985

[80] J. Weng, T. S. Huang & N. Ahuja, "3-D Motion Estimation, Understanding, and Prediction from Noisy Image Sequences," IEEE Trans. Pattern Anal. Machine Intell., Vol. PAM1-9, No.3, May 1987

[81] R. J. Crinon, "Contour-Based Motion Estimation," Tektronix Report No. ESL-DSPP031/87

BIOGRAPHICAL NOTE

The author was born 13 July 1961, in Shanghai, China. In 1979, she attended Tsinghua University, Beijing, China, where she received her Bachelor and Master of Electrical Engineering in 1984 and 1987, respectively.

In July 1987, the author began a teaching position at Tsinghua University. For the period June to October 1987, she worked at Beijing University of Aeronautics and Astronautics as a visiting researcher.

In October 1988 the author began her studies at the Oregon Graduate Center where she completed the requirements for the degree Master of Science in December 1989.

The author is leaving the Oregon Graduate Center to continue her education at the University of Washington, Seattle.