UNIT 4: VIDEO AND ANIMATION

Video is a combination of image and audio. It consists of a set of still images called frames displayed to the user one after another at a specific speed, known as the frame rate measured in number of frames per second (fps), If displayed fast enough our eye cannot distinguish the individual frames, but because of persistence of vision merges the individual frames with each other thereby creating an illusion of motion. The frame rate should range between 20 and 30 for perceiving smooth realistic motion. Audio is added and synchronized with the apparent movement of images.

1.1 VIDEO SIGNAL REPRESENTATION

In conventional black-and-white TV sets, the video signal is displayed using a CRT (Cathode Ray Tube). An electron beam carries corresponding pattern information, such as intensity in a viewed scene. To understand later reasoning behind data rates of motion video and computer-based animation, we focus on the description of their respective signals rather than specific camera or monitor technologies. We analyse the video signal coming from camera and the resulting pictures (using USA standards).

Video signal representation includes three aspects:

- Visual representation
- Transmission
- Digitalization

1.1.1 Visual representation

A central objective is to offer the viewer a sense of presence in the scene and of participation in the events portrayed. To meet this objective, the televised image should convey spatial and temporal content of the scene. Important measures are:

1. **Vertical Detail and Viewing Distance:**
   The geometry of the field occupied by the television image is based on the ratio of the picture width W to height H. It is called aspect ratio. The conventional aspect ratio is 4/3=1.33.
   The smallest detail that can be reproduced in the image is a pixel. Ideally, each detail of the scene would be reproduced by one pixel. Practically, however, some of the details in the scene inevitably fall between scanning lines, so that two lines are required for such picture elements. Thus, some vertical resolution is lost. Measurements of this effect show that only about 70% of the vertical detail is presented by the scanning lines. The ratio is known as the Kell factor; it applies irrespective of the manner of scanning, whether the lines follow each other sequentially (a progressive scan) or alternately (an interlaced scan).

2. **Horizontal Detail and Picture Width:**

The picture width chosen for conventional television service is 4/3*picture height. Using the aspect ratio, we can determine the horizontal field of view from the horizontal angle.

3. **Total Detail Content of the Image:**
   The vertical resolution is equal to the number of picture elements separately presented in the picture height, while the number of elements in the picture width is equal to the horizontal resolution times the aspect ratio. The product of the number of elements vertically and horizontally equals the total number of picture elements in the image.

4. **Perception of Depth:**
   In natural vision, perception of the third spatial dimension, depth, depends primarily on the angular separation of the images received by the two eyes of the viewer. In the fat image of television, a considerable degree of depth perception is inferred from the perspective appearance of the subject matter. Further, the choice of the focal length of lenses and changes in depth of focus in a camera influence the depth perception.

5. **Luminance and Chrominance:**
   Color vision is achieved through three signals, proportional to the relative intensities of Red, Green and Blue light (RGB) in each portion of the scene. The three signals are conveyed separately to the input terminals of the picture tube, so that the tube reproduces at each point the relative intensities of red, green and blue discerned by the camera.
   During the transmission of the signals from the camera to the receiver (display), a different division of signals in comparison to the RGB division is often used. The color encoding during transmission uses luminance and two chrominance signals.

6. **Temporal Aspects of Illumination:**
   Another property of human vision is the boundary of motion resolution. In contrast to continuous pressure waves of an acoustic signal, a discrete sequence of individual pictures can be perceived as a continuous sequence. This property is used in television and motion pictures, i.e., motion is the presentation of a rapid succession of slightly different still pictures (frames). Between frames, the light is cut off briefly. To represent visual reality, two conditions must be met. First, the rate of repetition of the images must be high enough to guarantee smooth motion from frame to frame. Second, the rate must be high enough so that the persistence of vision extends over the interval between flashes.

7. **Continuity of motion:**
   It is known that we perceive a continuous motion to happen at any frame rate faster than 15 frames per second. Video motion seems smooth and is achieved at only 30 frames per second, when filmed by a camera and not synthetically generated. Movies, however, at 24 frames/s. The new Show scan technology involves making and showing movies at 60 frames per second and on 70-millimeter films. This scheme produces a bigger picture, which therefore occupies a larger portion of the visual field, and produces much smoother motion.

There are several standards for motion video signals which determine the frame rate to achieve proper continuity of motion. The USA standard for motion video signals, **NTSC (National Television Systems Committee)** standard, specified the frame rate initially to 30 frames/s, but later changed it to 29.97 Hz to maintain the visual-aural carrier separation at precisely 4.5 MHz. NTSC scanning equipment presents images at the 24 Hz standard, but transposes them to the 29.97 Hz scanning rate. The European standard for motion video, PAL (Phase Alternating Line), adopted the repetition rate of 25 Hz, and the frame rate therefore is 25 frames/s.

8. **Flicker:**
   Through a slow motion, a periodic fluctuation of brightness perception, a flicker effect, arises. The marginal value to avoid flicker is at least 50 refresh cycles/s. To achieve continuous flicker-free motion, we need a relatively high refresh frequency. Movies, as well as television, apply some technical measures to work with lower motion frequencies.

9. **Temporal Aspect of Video Bandwidth:**
   An important factor to determine which video bandwidth to use to transmit motion video is its temporal specification. Temporal specification depends on the rate of the visual system to scan pixels, as well as on the human eye's scanning capabilities. For example, in a regular TV device, the time consumed in scanning lines and frames is measured in microseconds. In an HDTV (High Definition TV) device, however, a pixel can be scanned in less than a tenth of a millionth of a second.
   From the human visual perspective, the eye requires that a video frame be scanned every 1/25 second. This time is equivalent to the time during which a human eye does not see the flicker effect.

1.1.2 Transmission

Video signals are transmitted to receivers through a single television channel. The NTSC channel is shown in Figure below:

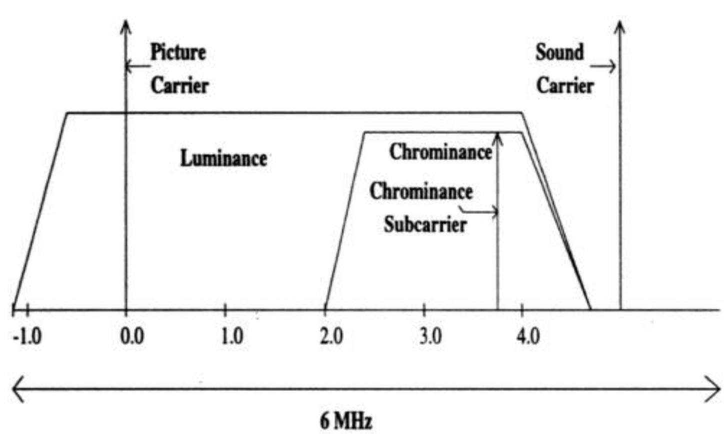

Figure 1: Bandwidth of the NTSC system

To encode color, a video signal is a composite of three signals. For transmission purposes, a video signal consists of **one luminance** and **two chrominance** signals. Luminance as different shades of light in grays while chroma are different hues (shade) of color.

In NTSC systems, the composite transmission of luminance and chrominance signals in a single channel is achieved by specifying the chrominance subcarrier to be an odd multiple of one-half of the line-scanning frequency. This causes the component frequencies of chrominance to be interleaved with those of luminance. The goal is to separate the two sets of components in the receiver and avoid interference between them prior to the recovery of the primary color signals for display.

Several approaches of color encoding are: RGB signal, composite signal

1.1.3 Digitalization

Before a picture or motion, video can be processed by a computer or transmitted over a computer network, it needs to be converted from analog to digital representation. In an ordinary sense, digitalization consists of sampling the gray (color) level in the picture at M x N array of points. Since the gray level at these points may take any value in a continuous range, for digital processing, the gray level must be quantized. By this we mean that we divide the range of gray levels into K intervals, and require the gray level at any point to take on only one of these values. For a picture reconstructed from quantized samples to be acceptable, it may be necessary to use 100 or more quantizing levels.

When samples are obtained by using an array of points or finite strings, a fine degree of quantization is very important for samples taken in regions of a picture where the gray (color) levels change slowly. The result of sampling and quantizing is a digital image (picture), at which point we have obtained a rectangular array of integer values representing pixels.

The next step in the creation of digital motion video is to digitize pictures in time and get a sequence of digital images per second that approximates analog motion video.

1.2 COMPUTER VIDEO FORMAT

The computer video format depends on the input and output devices for the motion video medium. Current video digitizers differ in **digital image (frame) resolution**, **quantization** and **frame rate** (frames/s).

NTSC video signal and after digitalization can achieve spatial resolution of 640 x 480 pixels, quantization of 8 bits/pixel (256 shades of gray) and a frame rate of 4 frames/second. The Sun Video digitizer from Sun Microsystems, on the other hand, captures NTSC video signal in the form of the RGB signal with frame resolution of 320 x 240 pixels, quantization of 8 bits/pixel, and a frame rate of 30 frames/second.

The output of the digitalized motion video depends on the display device. The most often used displays are raster displays, described in the previous chapter. A common raster display system architecture is shown in below:

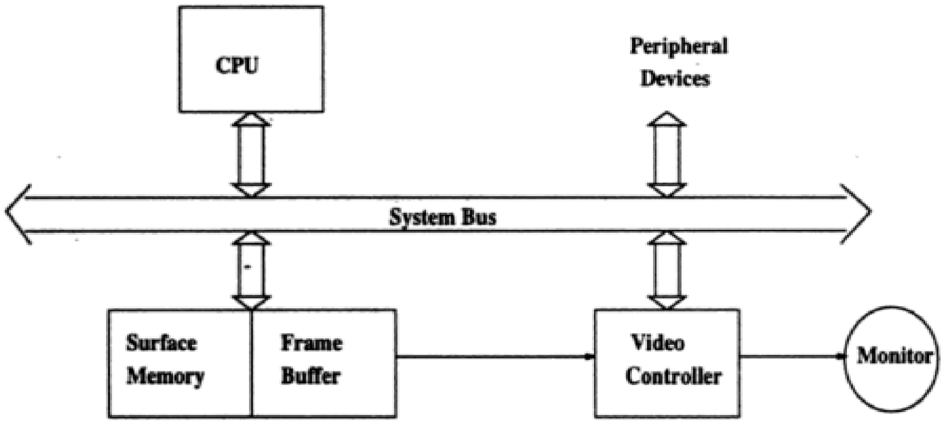

Figure 2: Display for video controller

The video controller displays the image stored in the frame buffer, accessing the memory through a separate access port as often as the raster scan rate dictates. The constant refresh of the display is its most important task. Because of the disturbing flicker effect, the video controller cycles through the frame buffer, one scan line at a time, typically 60 times/second. For presentation of different colors on the screen, the system works with a Color Look Up Table (CLUT or lut). At a certain time, a limited number of colors (n) is prepared for the whole picture. The set of n colors, used mostly, is chosen from a color space, consisting of m colors, where generally n << m.

Some of the computer video controller standards are:

- **CGA** (Color Graphics adapter): The first color monitor and graphics cards for PC computers. Capable of producing 16 colors at 160x200 pixels.
- **EGA** (Enhanced Graphics Adapter): an adapter that could display 16 colors with a screen resolution of 640x350 pixels.
- **VGA** (Video Graphics Adapter): Currently the base standard for PC video cards and monitors. True VGA supports 16 colors at 640x480 pixels or 256 colors at 320x200 pixels.
- **SVGA** (Super VGA): A SVGA card or monitor is capable of displaying more pixels (dots on the screen) and/or colors than basic VGA. For example, an SVGA graphics card may be able to display 16-bit color with a resolution of 800x600 pixels.
- **XGA** (Extended Graphics Array): A standard used on some IBM PS/2 models. XGA supports 256 colors at 1024x728 pixels, or 16-bit colors at 640x480 pixels.

1.3 COMPUTER BASED ANIMATION

Animation means giving life to any object in computer graphics. It has the power of injecting energy and emotions into the most seemingly inanimate objects. Computer-assisted animation and computer-generated animation are two categories of computer animation. It can be presented via film or video.

An animation covers all changes that have a visual effect. Visual effects can be of different nature. They might include time-varying positions (motion dynamics), shape, color, transparency, structure and texture of an object (update dynamics), and changes in lighting, camera position, orientation and focus.

A computer-based animation is an animation performed by a computer using graphical tools to provide visual effects. Processes of computer based animation are as follows:

1.3.1 Input Process

Before the computer can be used, drawings must be digitized because key frames, meaning frames in which the entities being animated are at extreme or characteristic positions, must be drawn. This can be done through optical scanning, tracing the drawings with a data tablet or producing the original drawings with a drawing program in the first place. The drawings may need to be post-processed (e.g., filtered) to clean up any glitches arising from the input process.

1.3.2 Composition Stage

The composition stage, in which foreground and background figures are combined to generate the individual frames for the final animation, can be performed with image-composition techniques. By placing several low-resolution frames of an animation in a rectangular array, a trail film (pencil test) can be generated using the pan-zoom feature available in some frame buffers. The frame buffer can take a particular portion of such an image (pan) and then enlarge it to fill the entire Screen. This process can be repeated on several frames of the animation stored in the single image. If it is done fast enough, it gives the effect of continuity. Since each frame of the animation is reduced to a very small part of the total image (1/25 or 1/36), and then expanded to fill the screen, the display device's resolution is effectively lowered.

1.3.3 Inbetween Process

The animation of movement from one position to another needs a composition of frames with intermediate positions (intermediate frames) inbetween the key frames. This is called the inbetween process. The process of inbetweening is performed in computer-based animation through interpolation. The system gets only the starting and ending positions.

Inbetweening also involves interpolating the shapes of objects in intermediate frames. Interpolation may be either **linear** or **spline.** They made a skeleton for a motion by choosing a polygonal arc describing the basic shape of a 2D figure (or portion of a figure) and a neighbourhood of this arc. Inbetweening is performed by interpolating the characteristics of the skeleton between the key frames. A similar technique can be developed for 3D, but generally interpolation between key frames is a difficult problem.

1.3.4 Changing Colors

For changing colors, computer-based animation uses CLUT (lut) in a frame buffer and the process of double buffering. The lut animation is generated by manipulating the lut. The simplest method is to cycle the colors in the lut, thus changing the colors of the various pieces of the image. Using lut animation is faster than sending an entire new pixmap to the frame buffer for each frame. Assuming 8 color bits per pixel in a 640 x 512 frame buffer, a single image contains 320 Kbytes of information. Transferring a new image to the frame buffer every 1/30 of a second requires a bandwidth of over 9 Mbytes per second. On the other hand, new values for the lut can be sent very rapidly, since luts are typically on the order of a few hundred to a few thousand bytes.

1.4 ANIMATION LANGUAGE

There are many different languages for describing animation, and new ones are constantly being developed. They fall into three categories:

1.4.1 Linear-list Notations

In linear-list notations for animation each event in the animation is described by a starting and ending frame number and an action that is to take place (event). The actions typically take parameters, so a statement such as

42, 53, B, ROTATE PALM'',1 ,30

Means "between frames 42 and 53, rotate the object called PALM about axis 1 by 30 degrees, determining the amount of rotation at each frame". Many other linear-list notations have been developed, and many are supersets of the basic linear-list idea. An example is Scefo (Scene Format), which also includes the notion of groups and object hierarchy and supports abstractions of changes (called actions) using higher-level programming language constructs.

1.4.2 General-purpose Languages

Another way to describe animation is to embed an animation capability within a general-purpose programming language. The values of variables in the language can be used as parameters to the routines, which perform the animation.

ASAS is an example of such a language. It is built on top of LISP, and its primitive entities include vectors, colors, polygons, solids, groups, points of view, subworlds and lights. ASAS also includes a wide range of geometric transformations that operate on objects.

1.4.3 Graphical Languages

One problem with textual languages is inability to visualize the action by looking at the script. If a real-time previewer for textual animation languages were available, this would not be a problem; unfortunately the production of real-time animation is still beyond the power of most computer hardware. Graphical animation languages describe animation in a more visual way. These languages are used for expressing, editing and comprehending the simultaneous changes taking place in an animation. The principal notion in such languages is substitution of a visual paradigm for a textual one. Rather than explicitly writing out descriptions of actions, the animator provides a picture of the action.

Examples of such systems and languages are GENESYS™, DIAL and S-Dynamics System.

1.5 METHODS OF CONTROLLING ANIMATION

Controlling animation is independent of the language used for describing it. Animation control mechanisms can employ different techniques.

1.5.1 Full Explicit Control

Explicit control is the simplest type of animation control. Here, the animator provides a description of everything that occurs in the animation, either by specifying simple changes, such as scaling, translation, and rotation, or by providing key frame information and interpolation methods to use between key frames. This interpolation may be given explicitly or (in an interactive system) by direct manipulation with a mouse, joystick, data glove or other input device. An example of this type of control is the BBOP system.

1.5.2 Procedural Control

Procedural control is based on communication between various objects to determine their properties. Procedural control is a significant part of several other control mechanisms. In particular, in physically-based systems, the position of one object may influence the motion of another (e.g., balls cannot pass through walls); in actor based systems, the individual actors may pass their positions to other actors to affect the other actors' behaviours.

1.5.3 Constraint-based Systems

Some objects in the physical world move in straight lines, but many objects move in a manner determined by other objects with which they are in contact, and this compound motion may not be linear at all. Such motion can be modelled by constraints. Specifying an animated sequence using constraints is often much easier to do than using explicit control. Systems using this type of control are Sutherland's Sketchpad or Borning's ThingLab. The extension of constraint-based animation systems to support a hierarchy of constraints and to provide motion where constraints are specified by the dynamics of physical bodies and structural characteristics of materials is a subject of active re- search.

1.5.4 Tracking Live Action

Trajectories of objects in the course of an animation can also be generated by tracking live action. Traditional animation uses rotoscoping. A film is made in which people/animals act out the parts of the characters in the animation, then animators draw over the film, enhancing the background and replacing the human actors with their animated equivalents.

Another live-action technique is to attach some sort of indictor to key points on a person's body. By tracking the positions of the indicators, one can get locations for corresponding key points in an animated model. An example of this sort of interaction mechanism is the data glove, which measures the position and orientation of the wearer's hand, as well as the flexion and hyperextension of each finger point.

1.5.5 Kinematics and Dynamics

Kinematics refers to the position and velocity of points. A kinematic description of a scene, for example, might say, "The cube is at the origin at time t =0. It moves with a constant acceleration in the direction (1, 1, 5) thereafter.

By contrast, dynamics takes into account the physical laws that govern kinematics (e.g. Newton's laws of motion for large bodies, the Euler-Lagrange equations for fluids, etc.). A particle moves with an acceleration proportional to the forces acting on it, and the proportionality constant is the mass of the particle. Thus, a dynamic description of a scene might be, "At time t =0 seconds, the cube is at position (0 meters, 100 meters, 0 meters). The cube has a mass of 100 grams. The force of gravity acts on the cube." Naturally, the result of a dynamic simulation of such a model is that the cube falls.

1.6 DISPLAY OF ANIMATION

To display animations with raster systems, animated objects (which may consist of graphical primitives such as lines, polygons, and so on) must be scan-converted into their pixmap in the frame buffer. To show a rotating object, we can scan-convert into the pixmap successive views from slightly different locations, one after another. This scan-conversion must be done at least 10 (preferably 15 to 20) times per second to give a reasonably Smooth effect; hence a new image must be created in no more than 100 milliseconds. From these 100 milliseconds, scan-converting should take only a small portion of time.

For example, if scan-converting of an object takes 75 milliseconds, only 25 milliseconds remain to erase and redraw the complete object on the display, which is not enough, and a distracting effect occurs. Double-buffering is used to avoid this problem. The frame buffer is divided into two images, each with half of the bits per pixel of the overall frame buffer. As an example, we describe the display of the rotation animation. Let us assume that the two halves of the pixmap are $image_0$ and $image_1$.

Load look-up table to display values as background color

Scan-convert object into $image_0$

Load look-up table to display only $image_0$

Repeat

Scan-convert object into $image_1$

Load look-up table to display only image

Rotate object data structure description

Scan-convert object into $image_0$

Load look-up table to display only $image_0$

Rotate object data structure description

Until (termination condition).

If rotating and scan-converting the object takes longer than 100 milliseconds, the animation is quite slow, but the transition from one image to the next appears to be instantaneous. Loading the look-up table typically takes less than one millisecond.

1.7 TRANSMISSION OF ANIMATION

As described above, animated objects may be represented symbolically using graphical objects or scan-converted pixmap images. Hence, the transmission of animation over computer networks may be performed using one of two approaches:

- The symbolic representation (e.g. circle) of animation objects (e.g. ball) is transmitted together with the operation commands (e.g. roll the ball) performed on the object, and at the receiver side the animation is displayed. In this case, the transmission time is short because the symbolic representation of an animated object is smaller in byte size than its pixmap representation, but the display time at the receiver takes longer because the scan converting operation has to be performed at the receiver side. In this approach, the transmission rate (bits/second or bytes/second) of animated objects depends
    1. On the size of the symbolic representation structure, where the animated object is encoded,
    2. On the size of the structure, where the operation command is encoded,
    3. On the number of animated objects and operation commands sent per second.

- The pixmap representation of the animated objects is transmitted and displayed on the receiver side. In this case, the transmission time is longer in comparison to the previous approach because of the size of the pixmap representation, but the display time is shorter because the scan-conversion of the animated objects is avoided at the receiver side. It is performed at the sender side where animation objects and operation commands are generated. In this approach, the transmission rate of the animation is equal to the size of the pixmap representation of an animated object (graphical image) multiplied by the number of graphical images per second.

THANK YOU!!!