Sampling Theory for Digital Video Acquisition: The Guide for the Perplexed User

Prof. Lenny Rudin[1,2,3]
Ping Yu[1]
Prof. Jean-Michel Morel[2]

Cognitech, Inc, Pasadena, California (1)
Ecole Normale Superieure de Cachan, France(2)
SMPTE Member (Hollywood Technical Section) (3)

Abstract:

Recently, the law enforcement community with professional interests in applications of image/video processing technology has been exposed to scientifically flawed assertions regarding the advantages and disadvantages of various hardware image acquisition devices (video digitizing cards). These assertions state a necessity of using SMPTE CCIR-601 standard when digitizing NTSC composite video signals from surveillance videotapes. In particular, it would imply that the pixel-sampling rate of 720*486 is absolutely required to capture all the available video information encoded in the composite video signal and recorded on (S)VHS videotapes. Fortunately, these statements can be directly analyzed within the strict mathematical context of Shannon's Sampling Theory, and shown to be wrong. Here we apply the classical Shannon-Nyquist results to the process of digitizing composite analog video from videotapes to dispel the theoretically unfounded, wrong assertions.

Introduction:

The field of image processing is a mature interdisciplinary science that spans the fields of electrical engineering, applied mathematics, and computer science. Over the last sixty years, this field accumulated fundamental results from the above scientific disciplines.
In particular, the field of signal processing has contributed to the basic understanding of the exact mathematical relationship between the world of continuous (analog) images and the digitized versions of these images, when collected by computer video acquisition systems. This theory is known as Sampling Theory, and is credited to Prof. Claude Shannon of MIT (the founder of Information Theory), though already proposed by harmonic analysts in the 1920s.

Recent advances in computing technology have put advanced digital image processing capabilities onto affordable personal computers, thus making it available to the public at large. Medical professionals, the banking industry, the security industry, manufacturing, and law enforcement are using image processing today. While the end applications are well understood by users, the casual users of the technology rarely comprehend the fundamental principles on which these applications are based. There are situations when certain fundamental scientific principles of image processing technology have to be well understood by the end user. Often, the layperson intuition about what is true and what is not true in the computing imaging practice flies in the face of what the mathematical theory tells us. And the questions of 'How many pixels' are enough to capture video information, and if 'The more pixels are always better' are perplexing to the casual user, particularly because the answer is not intuitive at all. Clearly, the field of forensic image (video) processing mandates a careful, enlightened use of this technology, since the outcome affects people's lives. Therefore, we will recount the classical Sampling Theory in the context of digitizing composite video sources.

1. Shannon's Sampling Theory and Practice:

Prof. Claude Elwood Shannon
The father of Information Theory
Born: 30 April 1916 in Gaylord, Michigan, USA
Died: 24 Feb 2001 in Medford, Massachusetts, USA

A. Theory [2,3]

We will consider an analog signal $x_a(t)$ and its corresponding digital sequence form $x(n)$, and find their relations in the sense that $x(n)$ is a true representation of $x_a(t)$.

The Fourier transform of the analog signal $x_a(t)$ is:

$$X_a(\Omega) = \int_{-\infty}^{\infty} x_a(t) e^{-j\Omega t} dt \quad (1)$$

$$x_a(t) = \tfrac{1}{2\pi} \int_{-\infty}^{\infty} X_a(\Omega) e^{j\Omega t} d\Omega \quad (2)$$

The digital sequence $x(n)$ is the sampling of $x_a(t)$ at sampling period T.

So we have $x(n) = x_a(nT) = \tfrac{1}{2\pi} \int_{-\infty}^{\infty} X_a(\Omega) e^{j\Omega nT} d\Omega$. (3)

The discrete Fourier transform of the discrete sampled signal $x(n)$ is

$$X(e^{j\omega}) = \sum_{n=-\infty}^{\infty} x(n) e^{-j\omega n}$$

Hence the inverse discrete Fourier transform of $X(e^{j\omega})$ is:

$$x(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} X(e^{j\omega})e^{j\omega n}d\omega \tag{4}$$

Equation (3) can be rearranged into the sum of integration of periods $\frac{2\pi}{T}$

$$x(n) = \frac{1}{2\pi}\sum_{m=-\infty}^{\infty}\int_{(2m-1)\pi/T}^{(2m+1)\pi/T} X_a(\Omega)e^{j\Omega nT}d\Omega = \frac{1}{2\pi}\sum_{m=-\infty}^{\infty}\int_{-\pi/T}^{\pi/T} X_a(\Omega+\tfrac{2\pi m}{T})e^{j\Omega nT}d\Omega\ e^{j2\pi mn}$$

With the substitution of $\Omega = \frac{\omega}{T}$ and exchange the integration order, we get

$$x(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left[\frac{1}{T}\sum_{m=-\infty}^{\infty} X_a(\tfrac{\omega}{T}+\tfrac{2\pi m}{T})\right]e^{j\omega n}d\omega \tag{5}$$

Compare (4) and (5) we can identify

$$X(e^{j\omega}) = \frac{1}{T}\sum_{m=-\infty}^{\infty} X_a(\tfrac{\omega}{T}+\tfrac{2\pi m}{T}) \qquad or \qquad X(e^{j\Omega T}) = \frac{1}{T}\sum_{m=-\infty}^{\infty} X_a(\Omega+\tfrac{2\pi m}{T}) \tag{6}$$

Equation (6) made a clear statement that the digital spectrum is the summation of the analog signal spectrum shifted by period $\frac{2\pi}{T}$. In order for the digital spectrum and analog spectrum to be identical, we must constrain the analog signal to be band limited, so that $X_a(\Omega)$ and $X_a(\Omega+\frac{2\pi}{T})$ do not overlap at the non-zero value region. Such a constraint is expressed as $X_a(\Omega) = 0 \quad for \ |\Omega| \geq \frac{\pi}{T}$, or the analog signal is limited in the band $|\Omega| < \frac{\pi}{T}$. We know $\frac{1}{T} = f_{sampling}$ is the sampling frequency and $\Omega = 2\pi f$ where f is the analog frequency.

We get $f_{sampling} > 2f$ (This critical sampling rate is called <u>*Nyquist*</u> frequency) (7)

Equation (7) gives us an important guideline of sampling the analog signal. The sampling rate (frequency) must be higher than twice the highest analog frequency.

Now assuming we meet the sampling constraint, we then have from (6)

$$X(e^{j\Omega T}) = \frac{1}{T}X_a(\Omega) \qquad -\pi/T \leq \Omega \leq \pi/T \tag{8}$$

We can construct the analog signal from its samples as follows: from (2) and (8)

$$x_a(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} X_a(\Omega)e^{j\Omega t}d\Omega = \frac{1}{2\pi}\int_{-\pi/T}^{\pi/T} X_a(\Omega)e^{j\Omega t}d\Omega = \frac{1}{2\pi}\int_{-\pi/T}^{\pi/T} TX(e^{j\Omega t})e^{j\Omega t}d\Omega$$

Since $X(e^{j\Omega t})$ is the discrete Fourier transform of $x(n)$, then $X(e^{j\Omega T}) = \sum_{m=-\infty}^{\infty} x(m)e^{-j\Omega Tm}$

Therefore $x_a(t) = \frac{1}{2\pi}\int_{-\pi/T}^{\pi/T} T\sum_{m=-\infty}^{\infty} x(m)e^{-j\Omega Tm}e^{j\Omega t}d\Omega = \sum_{m=-\infty}^{\infty} x(m)\left[\frac{T}{2\pi}\int_{-\pi/T}^{\pi/T} e^{-j\Omega Tm}e^{j\Omega(t-mT)}d\Omega\right]$

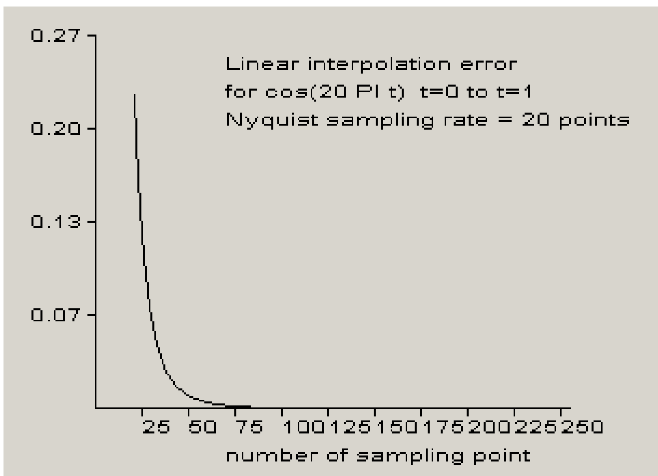
And we can reconstruct original analog signal:

$$x_a(t) = \sum_{m=-\infty}^{\infty} x(m)\frac{\sin[(\pi/T)(t-mT)]}{(\pi/T)(t-mT)} \tag{9}$$

B. Numerical Practice

The Sampling Theory tells us that as long as we choose the sampling rate within the Nyquist frequency, we can always reconstruct the original signal using equation (9). But in reality, we do not have an infinitely long sampling sequence $x(n)$. What we have is N points of sampling data. We should consider what effect this constraint implies to the digital system.

Assuming the original signal is $\cos(2\pi ft)$, our signal interval of interest is from $t = 0$ *to* $t = 1$. The sampling period is $T = 1/N$, with N being the number of sampling points in the interval. We calculated linear interpolation error and plotted in the following graph.



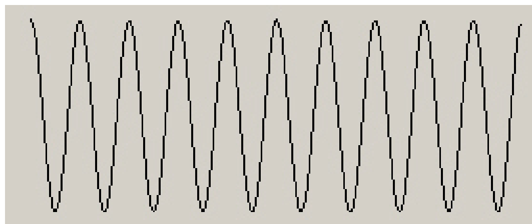
Plot 1

From the plot above, we can see that by slightly over sampling the original signal, the reconstruction error will reduce to zero very quickly (exponentially!) even for a very rough linear interpolation reconstruction. The numerical experiments showed similar behavior around the Nyquist frequency for other band-limited signals. This means, in simple terms, that significant over-sampling yields no further improvements!

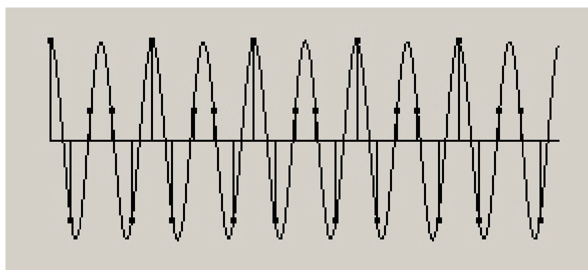
For the same signal, if we use zero padding in frequency domain, and inverse Fourier transform to obtain the interpolated signal at a smaller sampling interval, we can get the exact signal back. This means some signals can be reconstructed exactly, even at the Nyquist rate with a proper reconstruction procedure.

In general, for practical purposes, it is good to slightly over sample the signal above the Nyquist rate for the above stated reason of truncating the reconstruction series (9). However as the Error plot above shows, a few percent over-sampling suffices. The drawback of over-sampling is that we will end up with a larger amount of data.

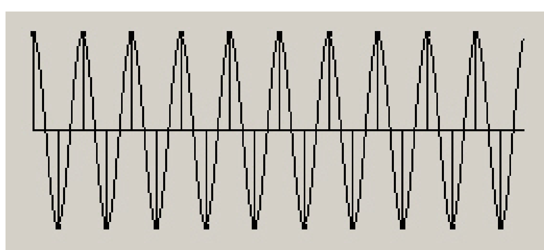
The following are examples of sampling above signal with different sampling rates and reconstruction by zero padding in Fourier transform domain.



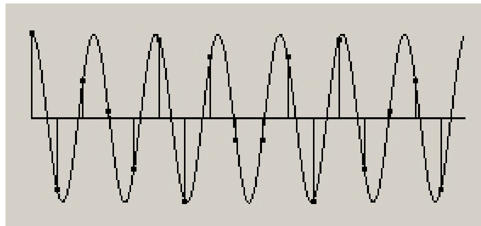
Original signal



Sampling over Nyquist rate (25 sample points)
(The reconstruction is good)



Sampling at Nyquist rate (20 sample points)
(The reconstruction is still good)



Aliasing: Sampling at below Nyquist rate (17 sample points)
(We see that the reconstruction does not reconstruct the original signal.)
Plot 2

2. Sampling Theory for Composite Video Sources:

There has been confusion recently, regarding the sampling rate of video images. Some systems sample a frame with 720x486, which is originally known as CCIR-601 Component Digital Standard, and other systems use a 640x480 rate, 'square pixel standard'. It has been wrongly claimed that the CCIR-601 standard is the "only right standard" to use when digitizing images from composite video sources such as NTSC, PAL, and SECAM. The origins of the CCIR-601 standard have *no relation* to improving digitizing composite video from e.g. (S) VHS videotapes. The main reason for SMPTE to develop the CCIR-601 standard was the widespread presence in the entertainment/broadcasting industry of component videotape recorders (DVTR's) recording 4:2:2 signals using the D1 standard, see [1] for complete treatment of D1 standard. A fascinating description of the Component Digital Standard CCIR-601 is in [6] where it is discussed why and how the 720*486 resolution was chosen by the SMPTE Committee. At the end of this section will make more remarks on the true origin of the 720*486 specific capturing rate. In what follows we show that the choice of either of the two sampling rate formats has no relevance to the claims of 'better image quality' when, for example, images are collected and recorded on VHS or SVHS video.

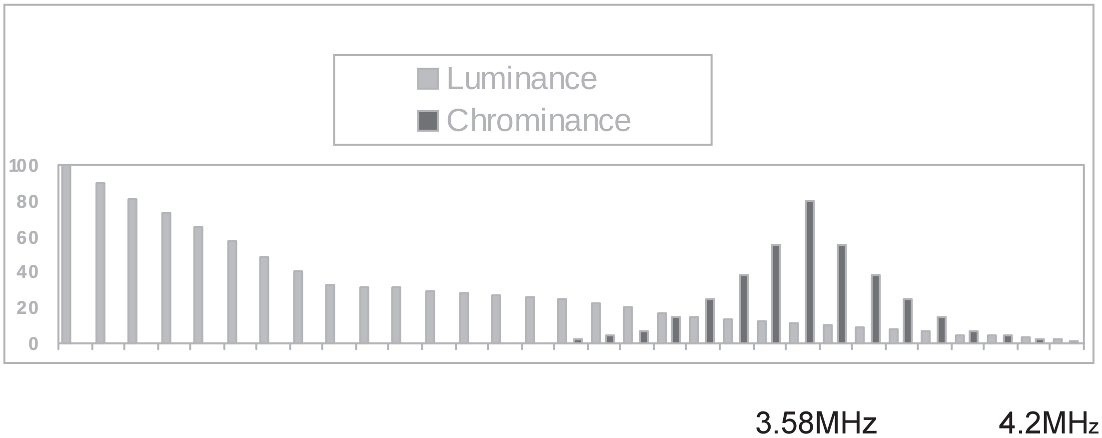
The legitimate question would be "How many pixels is enough to capture all available image data recorded on the composite videotape?" We shall apply the above-described Shannon's Sampling Theory.

In NTSC video, information is encoded in luminance and chrominance channels. There are 30 frames per second in the video signal or 60 fields per second. The scan line rate is therefore

(2.1)

For PAL video systems, the frame rate is 25 frames per second with 625 scan lines for each frame. Both NTSC and PAL have the same 15,750 Hz scan line rate.

Luminance signal and chrominance signal are mixed together by modulating the chrominance signal with a sub-carrier frequency at



Plot 3

The spectrum of the composite is shown on Plot3 above.

From the graph above, we can see that the maximum luminance frequency is 4.2MHz. Divide this number by the line rate 15750; we get the number of oscillations per scan line equal to 267. Using the Nyquist Rate (7), we find that 267x2 = 534 is the minimum sampling rate for a scan line. If we sample a scan line beyond 534 samples (pixels), we gain very little on most of the frequencies, and gain a little bit on higher frequencies. However, this gain exponentially decreases with over sampling percentage (plot 1). Accordingly to [1,5,7] an approximately 85% NTSC scan line contains image information, the rest devoted to horizontal blanking. Therefore 640*480 sampling rate represents approximately 20% - 30% over-sampling (depending if we count the non-image part of the frame that is approximately 15%). Therefore, any further over sampling is simply redundant.

The SVHS video cameras generate luminance signal with maximum frequency of 5MHz, which then gets recorded on the SVHS video tape. Thus similarly to the above computation, the corresponding Nyquist Rate is 635 samples per horizontal line, and if adjusted for the percentage of active image information as in above, the minimum number of pixels needed per scan line is 540. Thus sampling at 640 pixels per scan line represent about 16% over-sampling over Nyquist Rate, which is more than sufficient to avoid any information loss.

In an independent verification of our calculations in [7], p.20 in analysis by Prof. James Blinn, co-founder of science of computer graphics, shows that “a minimum number of image pixels needed per scan line for Y is 454” (i.e. 15% less than Nyquist Rate that we calculated! Due to the .85 proportion of active image information on the composite signal scan line).

Then why sample at 640*480 rather than much closer to the Nyquist rate? The answer is very simple yet misunderstood by many. When video is recorded there is ¾ ratio in the horizontal time rate to vertical time rate. That means that if equal number of pixels rows and columns is collected. e.g. 500*500, the resulting picture will be shrank by factor of ¾ in the horizontal direction. To avoid that the rate of 640*480 is chosen to reverse shrinkage in the displayed picture (640/480 = 4/3).

. Video can still be sampled at horizontal length of 720 pixels and vertical rate 486. However, no new information is gained by over sampling at this higher horizontal rate. But one has to be aware that when sampling a composite NTSC signal from (S)VHS videotape, a geometrical distortion will result. Namely, a circle on the videotape will be an ellipse with the ratio of 11% ellipticity, i.e. it will be 11% longer than taller. Some software packages hide this fact by rescaling the resulting digital image on the computer monitor. Typically a harmless scaling transformation is applied to fix this geometrical distortion [6]. So one cannot claim that the raw data was not processed. Fortunately this transform can be factored out when performing photogrammetry [4].

The following table is a summary of some NTSC and PAL systems in terms of their upper bound on the Nyquist scan line sampling rate. Make sure your video acquisition hardware complies with these requirements.

System	M/NTSC	G,H/PAL	I/PAL
Video Bandwidth (MHz)	4.2	5	5.5
Geographical Location	North America, Japan	Europe	Britain
Scanning	525/59.94	625/50	625/50
Sub-Carrier	3.58	3.58	4.43
Scan Line Rate	15,750	15,750	15,750
Oscillations/Sec	266.7	317.4	349
Nyquist Sampling Rate	534	635	698

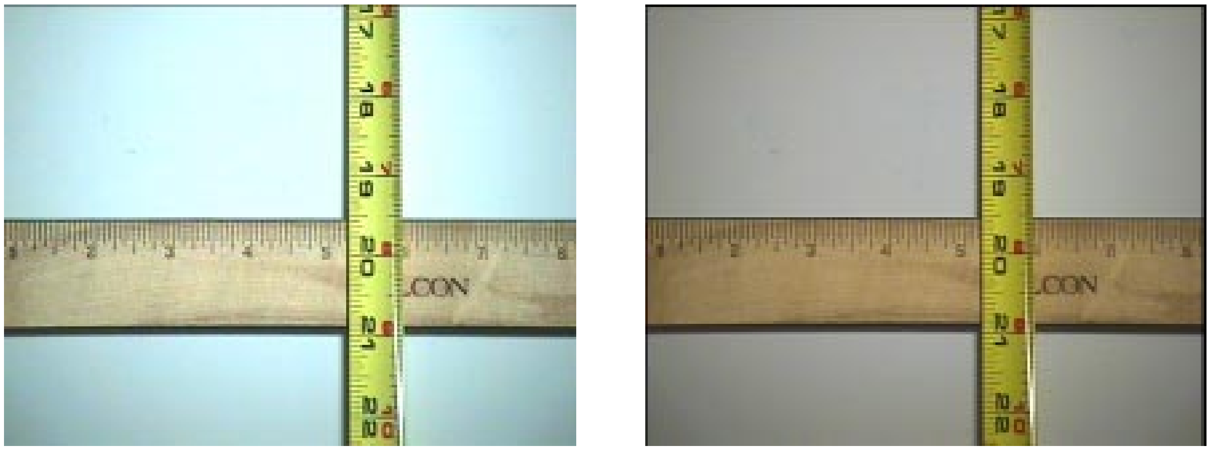
Table 1

As video is being displayed on the screen line by line, a small amount of time is needed to advance or retrace to the next line and from one frame to the next. These intervals are called blanking intervals. There are two types of blanking intervals -horizontal blanking and vertical blanking. The vertical blanking lies between the frames or fields of the image. The area needed for vertical blanking is approximately 8% of each frame period. If there are 525 scan lines in NTSC system, this gives us 483 (480) vertical lines comprising image information.

Note: S-Video (do not confuse with SVHS, which was considered above) signal has exactly the same Nyquist requirements as composite NTSC, since they only differ in the color/luminance encoding scheme, not in the spectral content [1].

To further demonstrate the above concepts of digitizing at the various rates above the Nyquist rate, we have split the same video output to two acquisition cards. The result is shown below. The pictures were rescaled to the same size by stretching, for comparison visualization.

Pic.1: Side by side comparison of the image capture at several above Nyquist rates



Capturing images at 640 x 480 resolution

Capturing images at 720 x 486 resolution (CC IR601)

Here the simultaneously captured video frames from **S-video** output show no essential resolution differences. The image composed of vertical and horizontal rulers was selected to do precise visual inspection of content differences, if any. One can notice that the CCIR card captures one-two pixel wide extra borders that normally are blank, i.e. have no visual content. This is hardly enough space to hide an image of a person with a gun, as has been asserted!

Pic 2: Side by side comparison of a natural image captured at two different, above Nyquist rates

640 x 480

720 x 486

It has been wrongly asserted that the non-CCIR-601 collected imagery will lose a large percentage of the information on the borders of the image, due to the insufficient sampling rate. To substantiate this false claim, examples like the picture below, usually would show a 640*480 frame within CCIR-601 frame, or side by side. The non-CCIR frame would be missing an entire person on the left or right fringe of the frame. As per the above analysis, this result is impossible. The examination of these 'examples' reveal that the only way it could be obtained is by first sampling at the 720*486 rate and then just cropping it in the center by a 640*480 sub window as shown below:

Cropping the CCIR-601 to 640*480

Of course if this is done, data will be missing on the borders. But in real case no cropping takes place with legitimate sampling hardware of 640*480 resolution. The authors of these false assertions also seem to confuse digitization by video acquisition cards with the so-called under/over scan feature of television monitors that leads to cropping. There is no relation here, the cards that the authors used in this paper both have under scan capability. Even more so, some of the 640*480 cards originally sample at 720*486 rate, the rescale it back to 640*480 to preserve the original geometry ratios. This is simply down sampling of the over sampled image, done in a careful way. In any case, the cropping 'proof' appears to be, at the very least, confusion and at the worst, a scientific deception.

One question is remaining: is why the SMPTE committee chose 720*486.
The answer is in the choice of the pixel clock at 13. 5 MHz to 'allow component digital recorders to operate similar rates, allow special effects, telecine machines operating at different rates etc.' [5]. These machines are not the common VCRs used by the law enforcement. They are DDR's like Sony D1 recorder, Digital Betacam, etc. A simple computation as in the prior NTSC derivation shows $858*f_h$ = 13.5 MHz as in (2.2). Then the only visible pixels are reduced to 720 [5].

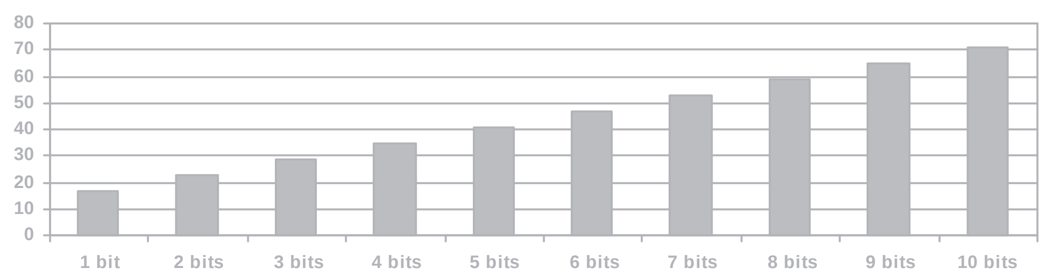
Since CCIR-601 has its origins in the entertainment/broadcasting industry, the video capturing hardware that supports this standard is way more expensive than the cards employed to perform tasks in industrial inspection/computer vision that are typically at 640*480. It appears that the manufacturers with entertainment/broadcasting roots would prefer to offer their more expensive hardware.

It is clearly shown in [6] that the CCIR-601 format is a choice preferred over 640*480 only from the point of ease of conversion between various composite and component video formats that include high end component technology recorders (used by news channels, and not by CCTV systems), and not because there is any loss of information when sampled at lower rates. And we agree with this assessment: the SMPTE standard is good overall convention, but not to the exclusion of other standards when acquiring video from (S)VHS tapes for the purposes of forensic video acquisition.

When the funding is an issue, the above described, more economical, 640*480 cards will capture good quality uncompressed video with no data loss.

3. Video Quantization and Noise:

The user should be aware of other theoretical and practical considerations when digitizing an image. Namely, most of the existing digitizers on the market offer eight bits per color per pixel. *Is there more data to be extracted with more bits per pixel?* A video signal whose amplitude takes a range of continuous value is quantized by assigning to each finite set of intervals of amplitude, a discrete number level. The effect of this quantization is equivalent to adding quantization noise to the ideal signal and reducing the signal to noise ratio (SNR) of the digital system. The theoretical SNR for an n-bit quantizer is:



Plot 4

From the Plot 4, we conclude that the more bits we quantize the signal, the higher SNR we gain. In practice, source video signal itself contains some amount of noise. A typical case is when a signal comes from worn out videotape. Since noise from the source and quantization noise are considered independent events, the total noise is their summation. One can argue that the more quantization bits we use, the better image we get. When quantization SNR is comparable with the source signal's SNR, the further increase of quantization bits does not significantly improve the resulting image. Our experiments with a ten-bit quantizer will be reported elsewhere. The preliminary data shows some improvement in feature extractions, particularly when the object of interest has to be contrast enhanced.

Conclusions:

The mathematical theory of image processing is a rich and useful field of scientific knowledge, constantly moving ahead due to the efforts of the scientific community. The field of forensic image/ video processing builds upon this strong scientific foundation. One shall be guided by this knowledge in order to elucidate various natural questions arising in the exploitation of image processing in general and in forensic video applications in particular. The beauty of mathematical theory is its absolute validity today as it will be hundred years from now. In this respect the above straightforward application of the classical Shannon Sampling Theory is a good example.

References:

1. 'A Technical Introduction to Digital Video' by Charles A. Poynton
2. 'Signal Analysis' by A. Papoulis
3. 'Digital Signal Processing' by Alan Oppenheim, Ronald Schafer
4. 'Geometrical Constraints in Forensic Video/Photogrammetry, Part 1' by L. Rudin et. al, Proceedings SPIE, 2002
5. 'The Art of Digital Video', by John Watkinson
6. 'The World of Digital Video' by Prof. James F. Blinn, Proceedings of IEEE Computer Graphics&Applications, September 1992, p.106-112
7. 'NTSC: Nice Technology, Super Color' Prof. James F. Blinn, Proceedings of IEEE Computer Graphics&Applications, March 1993, p.17 - 23