

HIGH QUALITY GEOMETRY DISTORTION TOOL FOR USE WITH LCD AND DLP PROJECTORS

Ronny Van Belle
Electronic Design Engineer
&
Bart Maximus
R&D Manager
Barco Projection Systems – Simulation Department
&
Philippe Vandenberg
Electronic Design Engineer
Barco Projection Products
&
Robert Clodfelter
Director of Engineering
EIS, Barco Group

Abstract

This paper describes a method of electronic geometry distortion correction for fixed resolution displays enabling the use of projection systems based on LCD or DMD technology in simulation applications where curved screens surrounding a user create a virtual environment.

In 1997 Barco introduced the first LCD projector with these capabilities but with no internal user interface. The use of an external PC was required and the distorted images suffered from a few artifacts such as aliasing effects and some sharpness and color detail loss.

Since then, advances in DSP technology made a lot of progress coupled with today's technological opportunities made it worthwhile to investigate how picture quality could be further improved.

The research described in this paper describes a method to improve the accuracy of the DSP circuits as well as newly developed DSP algorithms.

It also explains how a real time user interface for a geometric alignment may be implemented.

Introduction

Fixed resolution projection systems such as LCD and DLP(TM) based projectors have interesting properties when compared with CRT (or CRT addressed) projectors. Light output is much greater, the image appears sharper and set up is generally easier.

However, so far the typical application of these projectors is focussed on configurations in which the projector is orthogonally aligned (sometimes through the use of a folding mirror) before a flat screen. In such a situation, the projector usually can handle a number of different input formats. The different image formats are sampled and digitally remapped to the output resolution of the light valve.

This remapping or rescaling is often called scan conversion or pixel mapping, and is performed in an electronic module called a Pixel map Processor (PMP).

We distinguish between downscaling and upscaling. Downscaling is necessary to display a complete picture with a higher resolution than the display without losing information that was in the original picture while preserving the original sharpness as well as possible.

Conversely, upscaling is necessary in order to use the full display resolution and thus to obtain the maximal light output of the projector when driven by low-resolution sources.

Whenever the display configuration is different from the orthogonal alignment to the screen described above, the image on the screen will not be rectangular, and the necessary pixel mapping will be more complex than a simple shift or stretch in the horizontal and vertical directions.

Most PMPs can perform the more complicated non-uniform corrections in the horizontal directions on the screen (east-west correction), i.e. to compensate keystone correction when the projector is tilted vertically to the screen.

However, it is impossible to combine this with any type of correction in the vertical sense, the so-called north-south correction. In simulation applications this type of corrections is very often required, because of the widespread use of cylindrical, conical, toroidal and spherical screens, and the design of complete simulation systems where multiple projectors have to be put in "non-nominal" positions to integrate them nicely in the system.

The purpose of these non-conventional configurations is to increase the Virtual or Augmented Reality impression by surrounding the observer as much as possible.

Work in 1997 implemented an XGA capable PMP with both E-W and N-S mapping capability as shown in Fig 1.

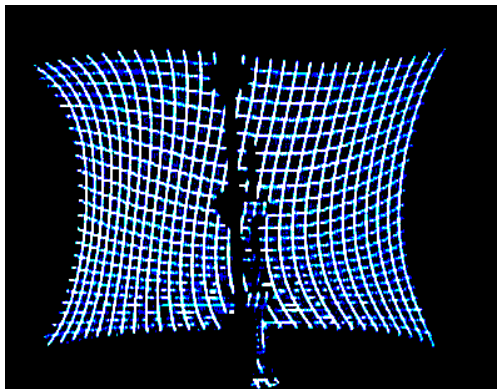


Fig. 1. Example of a photo taken from a geometrically distorted hatch pattern.

This design used dual ported memories combined with custom ASICs. Functionally, this design used a brute force approach of listing a repositioning value for each pixel in the output space. Image distortion then became a matter of creating a

large table of these values. The latency of this design was reduced to the minimum required for the maximum amount of vertical displacement in the output image space.

In 1999, Barco continued improving this concept and began the development of a geometry distortion tool that has a user-friendly interface for changing the geometry distortion parameters in real time and is capable of handling resolutions up to UXGA (1600 by 1280 pixels). It contains high precision calculations and newly developed filters to avoid aliasing effects.

The main text of this paper describes the improvements that were made by the new research done at Barco during the last year.

Main Text

Speed Limitations of Processing Hardware

There are two main reasons why the accuracy of the rescaling processing can never be infinite. First of all, there's the fact that these calculations must be done in real time, which means that with the current technology of the processing devices, even for the fastest gate arrays available today, the total number of calculations they can perform within a fixed period is limited.

Practically, if the number of bits of the video data is increased to gain picture quality, the number of useful calculation steps that can be done with these data will decrease along with the picture quality. A problem related to the previous explanation is that these arithmetic calculations usually need ripple carry chains.

Even with today's fastest programmable gate arrays it is impossible to build a 16 bit multiplier using carry chains that runs at the UXGA pixel rate of up to 200 MHz.

To achieve these high speeds, one may use a lot of pipelining but this is a device resource consuming solution with a potential power dissipation problem. The use of several parallel processing de-

vices can only partially overcome these speed limitations because with too many parallel devices the controller of the different processors becomes the bottleneck.

Random Access Limitations of Memory Devices

A second important reason why the bus width of the video data is limited is the technology of the memory devices that are necessary to store at least an area of the picture and also to perform a gamma function so that the interpolators can handle a linearized array of pixels. Without this gamma correction disturbing undulations of high frequency data can be seen on the screen. This means that for gamma corrected sources the video data has to be non-linearly transformed by a look-up table which is implemented in a random accessible memory that should run at the pixel rate.

Even the largest programmable devices contain insufficient memory to implement for example a 16-bit look-up table. Today's largest devices only contain about 40 Kbytes of embedded memory while a 16 bit look up table requires 128 kB. While it is easy to use twice the amount of memory for the storage of a 16-bit image instead of an 8-bit image, it is impossible to make all lookup tables 256 times as large to have 16 bit instead of 8 bit addresses.

This means the bus width of the video data has to be limited to achieve a real time speed performance and to have practical lookup table sizes. At this moment, fast memory devices exist with up to 36-bit width. This means 12 bits per color is about the greatest practical bus width today using lookup tables in programmable devices. However 12-bit precision is not enough to compensate for most gamma predistortions without the loss of image quality.

Required Bus Width to Perform Accurate Gamma Correction

The question arises how many bits are needed at the output of a look up table to perform a gamma correction for a given number of bits at the input without the loss of colors. For typical sources the value of

gamma is between 1.8 and 3.1 and for exponents larger than 1 this means that the attenuation of the signals is the strongest for the smallest video levels. To avoid the loss of colors, the contents of the look up table should be different for the lowest 3 addresses.

Suppose that the gamma function is truncated so that the exact value is replaced by the smallest integer larger than the exact value, then the contents of address 0 stays of course 0 and the result of address 1 is rounded to 1. In order to round the contents for address 2 to the value of 2, the result of the gamma function for the input of 2 must at least be larger than 1.

This can be expressed using the following formula:

$$(2 / 2^{Ni})^\gamma > 1 / 2^{No}$$

Which can be rewritten to answer the question above:

$$No > \gamma * (Ni - 1)$$

Where: Ni is the number of address bits of the look up table,

γ is the gamma exponent of the video source

No is the number of data bits required in the look up table.

Practically this formula means that for a true color video input having 8 bits per channel and a gamma factor of 3.1 a look up table is needed with a bus width of at least 22 bits to compensate for the gamma predistortion of that source. The interpolations should be performed with an accuracy of at least 22 bits and the result should be the address of a back end look up table of 4 megabytes which reapplies the gamma function on the video data.

Table 1. Required number of bits of the look-up table output as function of the gamma for true color inputs.

Gamma	Required number of bits
1	8
1.8	13
2	14
2.2	16
2.8	20
3.1	22
3.2	23

Table 1 gives an overview of the required number of bits for several typical gamma corrections to avoid the loss of color detail for a true color source.

Proposal of a Floating Point Number Format

As the above mentioned number of bits is not practical, another way to represent these large numbers caused by the gamma transform of the video values should be used.

During the design of the geometry distortion tool the researchers at Barco proposed a type of floating point representation which is well suited for today's practical electronic devices. These floating point numbers have 8 significant bits in a linear format accompanied by an exponent of 4 bits that represents the power of 2 by which the 8 bit number should be multiplied. To explain it another way, 8 significant bits are followed by up to 15 bits equal to zero (when the exponent is 15 which is the maximum value of a 4 bit number).

This means only 12 bits are needed to represent the values as accurate as a linear 23 bits equivalent without loss of video quality. An illustration of this floating-point format for a gamma correction look-up table is shown in table 2.

Table 2. Floating-point numbers stored in look-up table for a gamma of 3.1.

Hexadecimal 8-bit Input	Binary 23-bit Linear Output	Hexadecimal Floating-Point Output
0	000000000000000000000000	00e0
1	000000000000000000000001	01e0
2	000000000000000000000011	02e0
3	0000000000000000000001001	09e0
4	00000000000000000000010110	0Be1
5	000000000000000000000101011	2Be0
6	0000000000000000000001001011	4Be0
..
FE	11111011111010011010101	FBeF
FF	111111100000000000000000	FFeF

These 12 bits per color fit perfectly together in a 36 bit RAM that is an affordable standard device. With a linear equivalent of 23 bits this is sufficient to process true color sources with a gamma of up to 3.2

while maintaining all color details in the original signal.

Another reason why this floating-point representation of the video data is very effective is because of the nature of the distortion calculation itself. To calculate the output pixels for the display a weighted sum is calculated from an array of 4*4 source pixels.

A multiplier is needed to calculate the contribution of each source pixel to the display pixel. When using the floating point representation proposed above the width of these multipliers is limited to 8 bits instead of 23 bits. The effect of the exponent value is a simple shift register that consumes a negligible amount of logic.

Benefits Of Gamma Predistortion Using LCD or DMD based Projectors

Historically gamma predistortion was introduced to compensate for the non-linear behavior of CRT displays. Today most digital image generators such as the video graphics adapters used in personal computers have a built in possibility to match the PC-image to make it appear correctly on the connected monitor.

The same is true for simulation image generators where CRT based projectors are still frequently used to create a virtual environment. The sources are equipped with a gamma predistortion necessary for CRT displays.

When a LCD or DMD based projector is used in such a system, the gamma predistortion of the IG can usually be set to a linear response so that the gamma predistortion does not need to be removed again digitally at the front end of the projector. This saves hardware in the projector because each non-linear transform increases the required bus width as discussed above.

However, not only do CRT displays have a gamma response but so does the human eye. Practically this means that we can see much smaller differences in dark colors than we can see in bright tones. It is useful to quantize signals with a gamma predistortion of around 2.2 because in such a case all quantiza-

tion intervals will appear as equal steps for the eye.

In order to reduce the amount of distinguishable quantization effects for the human eye, it is best to maintain a gamma predistortion at the IG even when non-CRT based displays such as LCD projectors are used.

Primary Color Matching In Multiple Projector Set Ups

The need for high bus widths when applying non-linear transforms to the video data such as gamma corrections or the compensation for the non-linear transmission curve of the LCD panels is already discussed extensively.

However a high bus width is also necessary to perform primary color matching of different projectors. This correction is necessary wherever multiple projectors are displaying parts of one large image. Because of several optical tolerances the primary colors are never exactly the same.

Even with carefully selected optical components a small electronic correction remains necessary. This is realized by adding small fractions of the primary colors to the linear primary components. As for instance, the contribution of the green input to the red output is very small; it is necessary to calculate this matrix transform using a sufficient bus width.

The use of floating point numbers ensures that no data is lost by color matching different projectors electronically.

Improving Color Accuracy of The Display

The advantages of implementing the processing using floating point numbers have already been discussed. This paragraph elaborates on how the floating point numbers are converted back to a linear format. Some light valves are driven by analog signals and in such a case a digital to analog converter is inserted after the linearization of the digital data.

The most straight forward method to transform the floating point data to an analog signal is using the following steps :1) Convert the floating point number to a linear 23 bits data bus; 2) Truncate the 23-bit result by ignoring for instance the 13 least

significant bits; 3) Convert the linear 10-bit bus to an analog signal;

It can not be avoided to make this kind of quantizing error for individual samples of a video signal. However it is possible to achieve a zero average error for all color tones by modulating the video data before the bus is truncated with a kind of noise signal. When 23 bits have to be converted to 10 bits, this noise signal has a magnitude of 13 bits.

The noise signal can be generated in three ways: 1) By a pseudo random generator; 2) Diffusion of the errors made by a pixels neighbors or the same pixel in the previous field; 3) Adding an offset pattern which can be variable in time to the video data;

A more complex noise pattern can also be obtained by a combination of the above described techniques. In fact, a good combination of the different methods is necessary to avoid flicker effects caused by temporal modulation or disturbing patterns caused by spatial modulations. Interference of these dithering processes with the video data must be avoided.

A good dithering algorithm has to reproduce a 10 bit video bus with exactly the same average value as the original 23 bit it was derived from in an area as small as possible over a very short period of time to avoid visual artifacts.

Interpolation Algorithm

Not only the bus width influences the final picture quality of the processing, but also the algorithms that are used. To perform geometry distortion, an interpolator is used to resample the original source image.

The interpolator has to make at least a weighted sum of 2*2 pixels. These 4 pixels are the nearest neighbors of video value to be sampled. This method is called bilinear interpolation and is the most widespread processing technique because of its simplicity.

The interpolation process can be seen as a combination of a low pass filter and a sampler. The quality of that filter determines the picture quality. Frequencies higher than half the sampling fre-

quency should be suppressed as far as possible to avoid aliasing effects while frequencies below that should be preserved.

A better approach is bi-cubic interpolation which takes an area of 4*4 pixels into account to preserve more of the sharpness of the original picture. Bi-cubic interpolation is commonly used in high-end applications and gives better results than bilinear scaling, but still the remapping is far from the ideal sinc function.

Resampling an image using area based discrete cosine transforms can approach this ideal sinc function more closely. Using this technique it is possible to create filters with an infinite order with much better frequency responses than can be imagined with bicubic or bilinear scalers as shown in

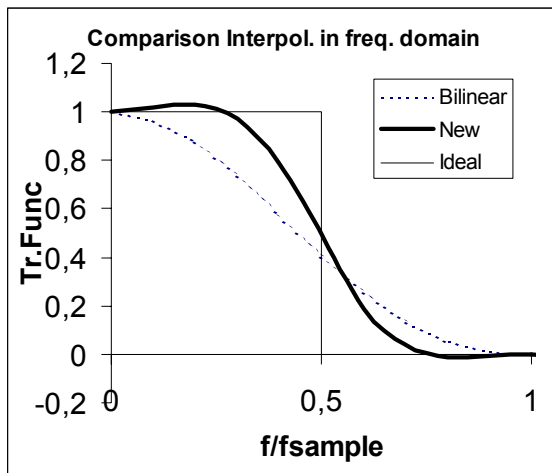


Fig. 2. Comparison of the frequency response of bilinear interpolation compared with the new technique based on discrete cosine transforms.

But even with an almost perfect static filter it is impossible to maintain the original image sharpness on every place of the display especially when the geometry distortion causes some areas to be resized by a factor close to 1:1. In such a case a pixel on pixel off pattern is displayed amplitude modulated dependent on the place. The size of these undulations depends on the scaling factor, which is area dependent in the case of geometry distortion.

These disturbing modulations are caused by the varying phase of the display pixels related to the original source pixels. This causes the average modulation trans-

fer function of the system to decrease which means that the image sharpness suffers from the rescaling process.

Improve Picture Sharpness

One way to overcome this is by changing the phase in a small area and keeping source pixels and display pixels as long as possible in phase in a larger area and repeat this process for the entire screen.

A disadvantage of such a sharpness boost is the apparent varying size of details in moving objects in the image.

Although the original picture sharpness is preserved, these motion artifacts do not enhance the appreciation of the images. In fact, the non-linear phase variations that sharpen the image introduce small repositioning errors.

So for static filters even with a good interpolator there's a trade off between sharpness and positional accuracy.

Improve Detail Size Accuracy

A better way to sharpen the image without introducing size changes of the details within moving objects is by changing the phase of the display pixels related to the source pixels adaptively.

The phase shift is done in areas that contain less detail than the average of the picture. This way the phase shift can not introduce artifacts in the details simply because there are no details in the neighborhood of the phase adjusted area.

A potential problem with this adaptive algorithm is the introduction of some intermodulation distortion caused by different parts of the images influencing each other. At this moment research is done at Barco about the behavior of different adaptive algorithms to manipulate the coefficients of the cosine transform used for the interpolation.

Repositioning Vectors

A lot of calculations are required to obtain the repositioning vector for each pixel of the display. This vector includes all necessary information for the interpolator to extract a display pixel from sev-

eral source pixels. It identifies which 16 pixels are required and what the contribution is of each for a specific display pixel. In other words it contains the read addresses of the frame memory and the coefficients for the interpolator. These values are derived from a unique set of coordinates for each pixel.

Vector Storage In a Frame Memory

One possibility is to store these coordinates for each display pixel in a memory with the depth of the display resolution. This is a simple hardware solution because an external computer can do all calculations of the repositioning vectors. Since the complete set of repositioning vectors describes entirely the geometry distortion one is also forced to have an external user interface for the set up of the projector.

This is not a user friendly solution because even with the most powerful PCs available today the calculations take a considerable amount of time and so does the transmission of the data from the PC to the projector because of the large amount of data.

Intermediate Vector Array

The only way to provide the user with a real time controllable interface is by reducing the number of parameters defining the geometry distortion.

Introducing an intermediate array of repositioning vectors solves this problem. This array has a relatively low resolution compared with the display but it is large enough to allow fine tuning of the geometry. For each area of 64*64 pixels the array contains an average repositioning vector. For an UXGA source with 1600*1280 pixels the array contains 25 columns and 20 rows of data which means that the entire geometry distortion is defined by 500 repositioning vectors.

Using the same area based cosine transform that is used for the video interpolation these 25*20 coordinates are scaled using a 24th order real time hardware filter to match the display resolution. But even the calculation of 500 coordinates is impossible in real time during adjustment.

For the first steps of a geometry calibration an even smaller array of 5*4 coordinates is used. This provides the user with a very fast responding, intuitive, coarse adjustment capability.

After the coarse alignment, most image parts are only displaced a few pixels from the designated position in any practical system geometry. All coarse reference points will be accurately displayed on the screen. Areas that are not yet displayed at the correct location can be modified using a fine calibration grid of 25 coordinates.

When this is still not sufficient for very complex screen types and projector configurations the user interface works with 80 positions.

For extremely critical set ups with several projectors each displaying a part of a large high resolution image of which parts are blended by at least two projectors the user can make very fine positioning corrections of a 32nd of the pixel size. In this case all 500 repositioning vectors of the intermediate array are used.

Repositioning Accuracy

The accuracy of the repositioning vectors is one 32nd of the pixel size. This means 5 bits are used to locate a display pixel between any two source pixels horizontally and the same is true for the vertical dimension.

When combined these 10 bits determine the weights of the surrounding pixels. This means that the coefficient that defines the contribution of each source pixel has a real accuracy of 10 bits.

The geometry distortion tool is designed theoretically for sources up to 2K*2K resolution which means 11 bits are required in both dimensions to address each pixel of the image individually.

The total size of the repositioning vector for each direction is 16 bit of which 11 bits are used to address the source pixels and 5 bits define the coefficients. Each of the 500 coordinates in the intermediate geometry distortion array has 32 bits of data. In that way

2Kbyte of data determines an entire geometry set up.

Without the use of an intermediate repositioning vector array remapped using the area based cosine transform these 2Kbyte of information would have been 8Mbyte.

The real time hardware interpolation of the coefficients is an essential element to achieve a user friendly interface while having a high repositioning accuracy.

Conclusion

The design of a high quality pixel map processing requires a series of trade-offs to maximize image quality and while controlling system cost. An approach is described that addresses the demands of both video non-linear gamma correction accuracy to 23 bits and video positioning to 1/32 of a pixel. Such accuracy is shown to be necessary to correctly map the input image. The selection of interpolation filter is also critical to the successful mapping process. Bilinear filters should not be used and bicubic filters are the minimum necessary to preserve image quality and detail. Using filters based on discrete cosine transforms even more details can be preserved.

All this processing must be performed in real time and without noticeable transport delay or quality loss. The improvement of the quality of the digital signal processing was realized thanks to new developments in the field of floating point processing as well as a new advanced method of interpolation that was proposed by the researchers at Barco.

Author Biography

In 1993 Ronny Van Belle joined the Barco company as an electronic design engineer in the lab of the projection systems department. There he first developed an ASIC that made a better color reproduction possible for the projectors by combining look up tables with several dithering algorithms in real time. Later, he studied the dynamic behavior of LCD panels and developed algorithms to compensate for different

artifacts introduced by LCD displays. He implemented compensations for transition speed, crosstalk, flicker and LCD panel non-uniformity in programmable logic. Features like true motion reproduction and true color reproduction are possible thanks to the hardware he implemented in programmable logic and ASICs. He also developed noise reduction algorithms, de-interlacing, contrast improvements based on histogram equalization and edge enhancement that is now used in all Barco LCD projectors. Currently Ronny Van Belle uses his experience in digital signal processing to design and implement the algorithms of the geometry distortion tool.

References

- Rogowitz, B. (1993). The human visual system : a guide for the display technologist. SID seminar lectures, ch. Session F-3,pp. F-3/39-F-3/56.
- Parker, J.A., Kenyon R.V., and Troxel, D.E. (1983). Comparison of interpolating methods for Image resampling. IEEE Transactions on Medical Imaging, Vol. MI-2, No. 1, March, 1983, pp. 31-39.
- Poynton, C. (1998). The rehabilitation of gamma. Proceedings of SPIE 3299, 1998, pp. 232-249