Technology and Play Pattern: Intel® Play™ Digital Movie Creator

Michael M. Conti, Smart Toy Lab (Consumer Products Division), Intel
Riley Jackson, Smart Toy Lab (Consumer Products Division), Intel
Henry Bruce, Smart Toy Lab (Consumer Products Division), Intel

Index words: Consumer Products, Extended PC, Smart Toys, Digital Video Processing, Audio

ABSTRACT

The flagship product for Intel Play for 2001 is the Digital Movie Creator (DMC), which is one of the many consumer products that came out of the Connected Products Division at Intel. The DMC is a hand-held digital video and still camera paired with easy-to-use, PC-based, movie-creation software (Figure 1). It is targeted at children, ages eight and up.

Figure 1: Digital Movie Creator

From early prototypes to final production, the DMC evolved through many stages of development: market research, usability studies, user interface design, and engineering feasibility studies. These stages helpto identify cost and implementation risks. We needed to ensure that we could deliver a $99USD product within budget and on schedule.

This paper provides a high-level outline of the technology drivers behind this unique movie-creation toy. These were the four drivers of the movie-making aspect of the product:

- image quality
- audio quality
- video frame rate
- audio video synchronization

We look specifically at how these technology drivers defined and were defined by the play pattern.

We also look at how end-user's expectations set some of the requirements for the technology drivers. And finally, we suggest what steps should be taken for future products that will ensure the proper compromises are made between technology and play patterns.

INTRODUCTION

The Intel® Play™ Digital Movie Creator (DMC) is a low-end digital alternative to using traditional film or camcorder video cameras. Instead of exposing film to light, the user is exposing light to an image sensor. Instead of playing back a film or videotape, the user is retrieving images from digital storage. It is a dual-mode digital camera that operates as both a digital camcorder and a digital still camera. It interfaces to a PC (Intel® Pentium® or Celeron® Processor MMX, 300MHz or better) via a USB connection. The camera may be operated in two modes: untethered from the PC (battery powered) or tethered, as a full-speed USB peripheral.

End-User Expectations

To determine an acceptable quality level is at best subjective. The determination of quality rests in large part on what the user's expectations are. Children can be a lot less demanding of quality than adults who are familiar with more high-end equipment. If the end-user is expecting quality standards commonly found in a Sony Hi8* camcorder when they use the DMC, they are bound to be disappointed. If the product delivers to their expectations for a children's toy, then the team has made the right compromises in choosing the technology drivers.

It was determined early on that the primary play pattern of the DMC is the capturing of content, either as movie clips or still images, and then playing them back on the PC. To deliver that experience, expectations for the product were that it *must* provide "good" quality in the following areas:

- image and audio
- video playback frame rate
- sychronized audio video

TECHNOLOGY DRIVERS

The five technology drivers to be discussed here are image quality, frame rate, audio quality, video-audio synchronization, and storage. Many of the technology drivers are dependent upon what choices the team made in their selection of hardware components.

Image Quality

Image quality can be affected by many variables, including lens selection, frame rate, and insufficient lighting. Poor design or a lack of attention to any one of these variables can result in poor images.

The toy needs an image sensor to capture snapshots and video clips. There are two flavors of image sensor: CCD and CMOS devices. CCD sensors have better performance; this is particularly noticeable in home lighting. The picture is sharper and less grainy and the colors stronger. However, CCD sensors are more expensive than CMOS image sensors. Due to the cost constraints, a CMOS sensor was used for the DMC and therefore image quality was sacrificed.

F-Stop and Lens Selection

Another variable in the image quality equation is the lens. A lens assembly is used to focus light onto the image sensor. The assembly is made up of a number of lens elements, which are made of either plastic or glass. Glass lens elements perform better, mainly because they let through more light and give a crisper image. However, plastic lens elements are considerably less expensive. Lens assemblies can be comprised of glass and/or plastic elements. A combination of element types is called a hybrid lens.

Also, the lens focus is fixed, as there is no way of checking focus when the DMC is untethered. Our decision to fix the lens focus was validated by the observation that children are not familiar with manual focus. Most children are familiar with simple "point-and-shoot" cameras.

Once again, a compromise was made due to cost, and the team decided to use a plastic lens. An f2.8 lens was used to ensure that the camera has a large depth of field. This was deemed to be more important than the greater light sensitivity offered by a lower f number lens. The f numbers indicate the size of the aperture (opening) relative to the focal length (distance from camera to subject). f numbers are calculated by dividing the focal length of the lens by the effective diameter of the aperture, e.g., 55mm lens, effective aperture 5mm = relative f11. Focus group feedback showed us it was better to make the camera work well in a variety of light settings with some loss of image quality.

This choice of the f stop is one of the compromises that came out of the play pattern usage model and which supported the secondary play pattern, stop-motion animation.

Stop-motion animation requires the user's explicit need to be close to the target that is being animated. Stop- motion animation involves the end-user manipulating an object in close range of the camera and taking frame-by- frame snap shots. The single frames are then combined into a single video (.avi) file.

Frame Rate

Frame rate isn't normally an issue when shooting in the traditional formats as video is fixed at 30 fps (frames per second) and film is 24 fps.

We looked at various frame rates with the DMC; both untethered and tethered operations, trying to decide from a user's standpoint what would be the best frame rate. It is important to note that in the digital format, a higher frame rate requires a faster processor and more storage space. Higher video frame rates would result in less available total record time (untethered). The team chose a lower end frame rate of 10 fps to provide the most flexibility to the product's play pattern.

Video Processing and Compression

Converting the raw data that comes from the image sensor into images requires complex Digital Signal- Processing (DSP) algorithms. Typically there are two forms of ASIC architecture that achieve this task: a programmable DSP core or hard-coded logic. The former is more flexible but more expensive, and it is used in products such as the Kodak mc3*. For the DMC, a "hard-coded" ASIC was used. An example of the compromise made here was that the ASIC had a bug that reduced the dynamic range of the processed image. If this had been a programmable device, the issue could have been corrected, but as the functionality was hard coded, this option was not possible.

To maximize the untethered recording time, the video has to be compressed. There are a number of algorithms suitable for this, and selecting one is usually a tradeoff between complexity, compressed image quality, and compression ratio. For the DMC, the Joint Photographic Experts Group (JPEG) compression algorithm was used, as this algorithm can be implemented as logic, resulting in lower ASIC cost. While the JPEG is more typically used for still image compression, for video each frame is JPEG compressed. This is termed Motion JPEG or MJPEG. Moving Pictures Experts Group (MPEG) is another video compression technique. It also compresses the difference between two subsequent frames (inter-frame compression), which results in a higher compression ratio at the cost of increased implementation complexity and cost.

Resolution

Although the image sensor has 352x288 pixels (known as CIF or Common Interchange Format), we used the central 320x240 Region Of Interest (ROI), as it allowed us to record 30% longer recording time without severely impacting overall image quality.

Audio Quality

Audio capture plays an important role in the creation of any movie. The DMC's play pattern supports this activity; it has a microphone built into the housing. While people can tolerate low-quality video, poor audio is much more noticeable and distracting. Thus, a significant effort was made to keep audio quality high. Microphone placement, digitization, and audio compression were areas that affected audio quality.

Microphone

An omni-directional microphone was used. Although a directional microphone is more desirable, it would have been too expensive. The microphone was positioned in the housing to ensure a maximum range of six feet, with minimal pick-up of handling noise and seismic rumble.

*Other names and brands may be claimed as the property of others.

Digitization

An AC97 compliant codec was used to digitize the signal from the microphone. Although this was "overkill" from a technical perspective, the ASIC has an AC97 input port, so this was a simple solution to implement.

The audio gain was set at the device's highest gain setting and stored in the internal SRAM. While it would have been better to provide an automatic gain control, this wasn't possible with the current design. Instead, we realized the user's desire to capture audio from at least four feet away and supported that with the high setting. When watching a movie, you notice the sound of someone's voice doesn't necessarily rise and fall based upon his or her distance from the camera, as this would be distracting. However, for the DMC, we had to compromise: the closer you are to the DMC, the louder you will sound and the further away you are, the softer you will sound.

Audio Compression

As for video, to maximize the video clip record time, the audio data should be compressed. Adaptive Differential Pulse Code Modulation (APDCM) compression was used for the normal setting. For the high-quality setting, the microcontroller executing the firmware did not have enough power to compress the audio, so the data were left in Pulse Code Modulation (PCM) format.

Video Audio Synchronization

As we know, a movie that is out-of-sync is pretty distracting. Some foreign movies, which are over-dubbed in English, come to mind: the action of the lips doesn't follow the words being heard.

In filmmaking, sound is usually recorded separately with specific sync points being generated (the reason for the "clap" board) and the end-result is synchronized in post-production. This final playback to film is generally done with optical audio tracks or magnetic tracks that run along the filmstrip. For the DMC, the end result needed to be the same as in film, i.e., the picture and audio had to be in sync, but the process of getting those two elements captured and played back wasn't the same.

Given that the application needed homogenous audio of 22KHz 16-bit mono PCM for simplified editing, the driver components responsible for download of audio from the camera to the PC up-sampled the audio stream. As mentioned above, the audio stored on the camera was in two different formats depending on the quality setting used when the audio/video was captured and stored. If the camera was in normal quality/resolution mode, then the audio was stored uncompressed at a 12KHz sample rate. If the normal quality/resolution setting was used, the audio was stored with a custom form of DPCM compression and sampled at 8KHz. Thus the host software took this into account when doing the sample rate conversion to the desired audio format prior to the creation of the final .avi file

Another benefit of allowing a quality choice was in the extension of the movie clip time: lower quality equals longer record times.

Integrating post capture sound in a movie clip happens a couple of different ways. In one scenario, the end-user captures someone singing Happy Birthday and then hears and sees it at her computer later. In another scenario, she could decide that the person singing Happy Birthday was doing a terrible job, so she mutes that voice and re-records over the same scene with her own. In yet another scenario, she could decide that the singing is fine but that it needs some background music. These opportunities to manipulate the post capture sound expanded the play pattern.

Storage

The camera requires some form of storage for snapshots and video clips captured when untethered. The cost of storage would limit us to two technologies: Synchronous Dynamic Random Access Memory (SDRAM) or FLASH (a marketing term of fast programmable EEPROM, or Electrically Erasable Programmable Read Only Memory). At the time of printing, SDRAM was one-fifth of the price of Intel StrataFlash®, so it was the chosen technology.

SDRAM is the most common type of computer memory; also known as D-RAM or DRAM. It usually uses one transistor and a capacitor to represent a bit. The capacitors must be energized hundreds of times per second in order to maintain the charges.

Unlike Flash, SDRAM needs an amount of current in order to retain its memory when "off," so a compromise was made to provide an off-on switch through button presses. In other words, the user had to first press a button to wake up the camera after it had gone to sleep. Then he had to press the button again to initiate the desired action, taking a snap shot or shooting video.

Video frames were captured to a temporary SDRAM buffer in CIF or QCIF resolution in YUV420 format. A more detailed discussion on YUV is presented in the hardware architecture section below.

Snapshots are video clips stored in SDRAM by using a simple filing system. A 64-byte header block precedes each item. For a snapshot, this is followed by the JPEG data, which is written to SDRAM directly by the JPEG compressor. For a video clip, the header is followed by interleaved audio and video data. The avi data comprises video frames with associated audio samples. Each frame is preceded by a 16-byte frame header, and is followed by the JPEG frame and then the audio samples that were captured during that frame time (copied from the internal SRAM). Due to restrictions of the filing system, no single data "block" can be larger than 32kbytes. This only affects 320x240 pixel JPEG snapshots, which could be of higher quality without this restriction. The frame header effectively acts as a time stamp so that audio and video data are synchronized.

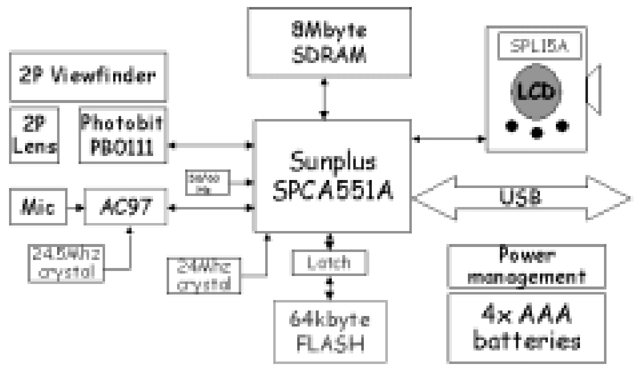

Figure 2: Hardware architecture

HARDWARE ARCHITECTURE

By discussing the technology drivers for component selection, we have almost covered the hardware architecture itself.

The largest, and most expensive component is the Application Specific Integrated Circuit (ASIC). Think of the ASIC as a chip that is custom designed for a specific application rather than a general-purpose chip such as a microprocessor. ASICs are "hardwired" to do a specific job and do not incur the overhead of fetching and interpreting stored instructions as in a computer. An ASIC chip performs an electronic operation as fast as it is possible to do so, providing, of course, that the circuit design is efficiently architected.

Some ASICs do the processing and compression of YUV data to JPEG data in firmware, but they need expensive processors or they are too slow. Doing the work in hardware makes the ASIC cheap, but you have to "approve" the processor blocks, as you can't change them.

The ASIC also interfaces to the rest of the components (e.g., image sensor, audio codec, SDRAM, etc.). The Sunplus SPCA551A was chosen as it met most of the play pattern requirements for a very competitive price. The

---

Intel StrataFlash is a registered trademark of Intel Corporation or its subsidiaries in the United States and other countries.

ASIC's central role in the hardware architecture is shown in Figure 2. An embedded 8052 microcontroller core executes the firmware out of external FLASH program's memory. This microcontroller is a minor Achilles heel for the ASIC: it was not powerful enough to handle the video data and compress the audio in the high-resolution mode.

The ASIC picked by the team utilized a Complementary Metal Oxide Semiconductor (CMOS) image sensor. Note that most video camcorders use a Charge Coupled Device (CCD) rather than CMOS. The ASIC colorizes the images, compresses the data using JPEG, and then stores them to memory or uploads them to the PC. To maximize video clip time (untethered) SIF (320x240 pixels) and QCIF (160x120 pixels) image sizes are used.

Although the CMOS sensor produces 352x288 (CIF), we used 320x240 (SIF) as it allowed us to record 30% more data into the camera's 8 megs of Synchronous Dynamic Random Access Memory (SDRAM).

Firmware

The complexity of the firmware was driven by the camera operations in the untethered mode. For instance, the whole system goes to sleep when not in use. It was a major firmware challenge to wake up, adjust exposure and white balance, and take a snapshot all in under two seconds (see Figure 2).

In the tethered mode, the firmware was primarily concerned about keeping the video stream going and tracking button presses.

SOFTWARE ARCHITECTURE

Figure 3 describes the breakdown of various software components that make up the DMC application and software stack.

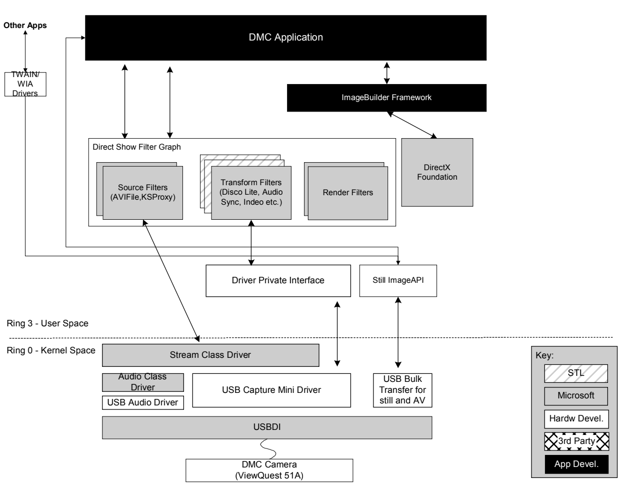

Figure 3: Software architecture overview

The application was responsible for creating and configuring the various DirectShow filter graphs for real-time capture and processing of audio/video data. Custom interfaces were used to interface with driver extensions implemented for specific DMC hardware that went beyond the scope of a typical camera device. A custom playback engine was developed for layer movie playback.

DirectShow*

The workhorse of the live video system was the Microsoft DirectShow* filter graph module. The filter graph was responsible for obtaining the audio/video stream from the DMC device and enhancing/converting each frame and then rendering it. This was accomplished by passing the video stream through several downstream filters in the filter graph. DirectShow provided an extremely flexible architecture in which filters are "plugged" into the graph to provide unique and specific behaviors.

The application interacted and controlled the DMC USB device via DirectShow's Capture Filter interfaces. A custom interface was exposed to the application, if needed, to allow it access and control over extended camera features.

Figure 4 below, describes the main filter graph that was configured by the application for preview and capture of video and snapshots. The camera provided I420 format video at CIF resolution (320x240). This video stream was then directed downstream into the AVI Decompressor, which produced RGB24. The Image Interpolation (Scale Image) filter was used to scale to video stream to a larger resolution for live preview at a larger size (1.3xSIF).

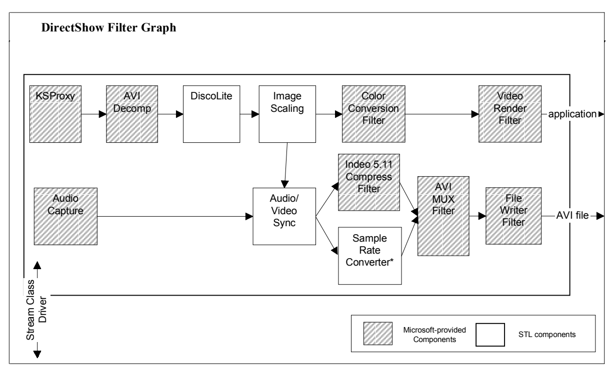

Figure 4: DMC "Live" Filter Graph

The camera also provides MS-class compliant audio. The in-camera USB microphone appears to DirectShow and

---

*Other brands and names may be claimed as the property of others.

Indeo is a registered trademark of Intel Corporation or its subsidiaries in the United States and other countries.

the application as a standard audio capture device. The audio stream sample rate, bit depth, and channels are converted to the desired format by the Audio Sample Rate Converter filter.

Compression

To conserve hard disk space, all AVI files were compressed using Indeo® 5.11. Video compression was done in the capture stream of the graph via the Indeo 5 real-time compressor.

Direct Show Filter Detail

Below are descriptions of the individual filters that comprise the various filter graphs.

KSProxy* Source Filter

The KSProxy source filter was provided by Microsoft and provided a generic interface to streaming video capture devices. It supported interfaces allowing access to standard camera properties and control (brightness, contrast, saturation, hue and video format, etc.).

Audio/Video Sync Filter & Mux

The audio/video filter used in conjunction with the Microsoft DirectX* 8 (DX8) MUX filter provide a reasonably reliable, single synchronized stream from the audio and video input sources. The audio/video sync filter's main task was to ensure the streams' start and stop points where synchronized. The MUX automatically drops and inserts video frames based upon the audio master stream and specified frame rate.

Image Interpolation Filter

The image interpolation filter was used to scale the video frames as they are sent downstream. This is a standard transform filter. This filter uses simple bilinear interpolation. Use of this filter provides a more dependable mechanism for image scaling when compared to the Microsoft video renderer filter, which can be video card dependent.

DiscoLite* Transform Filter

The DiscoLite transform filter first acts as a pin splitter. It takes the capture output pin from the KSProxy filter and splits it, thus allowing the application access to the two video output pins: capture and preview. These pins are then being controlled in a "gated" manner. If the gate is raised then the video stream flows downstream. If the gate is closed, video is not sent downstream. This provides the application considerable flexibility when controlling the filter graph when it is implementing features such as video capture. All pins on the DiscoLite filter (input, preview, capture) handled only RGB24 as a video subtype.

RESULTS

DMC's hardware and software architecture as described show that it is possible to engineer a product within the constraints of its technology drivers, and still provide a movie-creation experience. However, did DMC deliver to the user's expectations? Probably not if that person was expecting something similar to a home video camera. And to be honest, the quality just isn't there in terms of image and audio quality. But as a toy, it does deliver. The digital format holds much promise at the higher end products, but to justify a $99 USD price point, compromises were made in the selection and implementation of the drivers. These compromises were as follows:

- The image quality was lowered by the choice of a 2.8 f, which provides a wider variety of lighting situations during play.
- Audio-video synchronization is not ideal, but is acceptable to the end-user without adding to overhead performance on application side.
- Audio quality is acceptable for normal play pattern distance, 6 feet or less, but lacks auto gain controls and frequency controls for low-end sounds.
- There is a low frame rate, but it provides maximum ability for storage on the device.

At the end of the day, image and audio quality become less important. Children enjoyed the accessibility of the USB-enabled product and also seeing the immediate results of their work. Other movie-creation features not discussed here, such as sprite animations and painting on the video, also added value.

DISCUSSION

It is important to note that technology drivers enable the play pattern but don't make the toy. That experience comes from the end-user's experience with everything else, from the form factor to the other movie-creation activities. All of these items have equal importance in the final product. While Intel doesn't lack in its engineering ability, it must not lose sight of the big picture: to deliver to, and exceed, the consumer's expectations.

Here are some things that worked when designing the Digital Movie Creator (DMC).

- Evaluate the consumer's expectations with the product that is being built.

*Other brands and names may be claimed as the property of others.

For instance, we used focus group testing to determine if four minutes of untethered capture time was enough. A previous year's product was only going to use eight seconds!

- Make sure that the product has features that go beyond the primary feature–capturing movie clips.

  For the DMC, a strong secondary feature was the ability to do stop-motion animation and the fact that it had an array of creation tools.

- Observe the end-users interaction with the product.

  Do usability studies, and listen to the children. When image quality was an issue, it was generally thought to be the fault of the computer not being fast enough rather than the toy itself.

CONCLUSION

Technology is only one of the ingredients that make a consumer product great. How that technology is applied to an everyday task or activity and its relationship to a desired play pattern, makes a toy entertaining, fun, and unique. The Digital Movie Creator is a good example of a toy that takes the technologies of video capture and editing and applies them to storytelling, thereby creating a whole new category of play.

ACKNOWLEDGMENTS

Thanks to Jim Okuley (toy designer) for being the play pattern guru and everyone else on the DMC team who helped make this product a reality.

AUTHORS' BIOGRAPHIES

Michael M. Conti has been with Intel at the Smart Toy Lab since 1999, and is currently a program manager with CPD. Michael has 15 years of production management experience, from movie making to computer games. Michael has a B.A. degree in English. His e-mail address is Michael.M.Conti@intel.com

Riley Jackson is a Sr. Software Engineer and has been with Intel since 1991. He has been a software architect/developer at the Smart Toy Lab for two years. Past assignments include Intel Architecture Labs, Pro Share, and the Development Tools Org. He has 20 years experience in application and component design and development. His e-mail is riley.jackson@intel.com

Henry Bruce is a hardware engineer who has been with Intel at the Smart Toy Lab for a year. Before Intel, Henry worked for VLSI Vision Ltd. (VVL) in Edinburgh, latterly acquired by STMicroelectronics. There, as software development manager, he developed expertise in video processing and device drivers. He received his B.S. degree in Electronic Engineering in 1985 and his Ph.D. degree in Image Processing in 1993, both from the University of Edinburgh in Scotland. His e-mail address is henry.bruce@intel.com.