Investigations on Bits on Film high speed scanning device

Ulrich Klocke, Joachim Anders, Daniel Steiger, Andreas Hofmann; Fraunhofer IPM; Freiburg, Germany

Abstract

Whereas the recording of digital data on film for long term storage can be realized with high speed the reading of those data by scanning devices still needs significant more time. Additionally to that it is crucial to realize a fast scanning of the recorded data for verification of the integrity of the data. This is actually the bottleneck in realizing a fast archiving production chain "bits on film". We describe here a new approach for a high speed scanning system, where a new frame grabber based electronics to generate a fast data reduction of the scanned data is used. The amount of data generated by a high resolution scanner with four times oversampling rate is enormous. In order to read the "bits on film" (information data) with a line-sampling rate of 100kHz the high speed scanner has to acquire data with a data rate of about 1,6GByte/s. The challenge is to get the data amount reduced significantly already in the camera device without losing information. Therefore the region of information has to be separated and the data reduction has to be done geometrically and in bit depth. We have built up a test scanning device to verify the principal usage and functionality of this approach. We calculated that a data reduction rate of about 40 can be achieved. Although the focus of the test is on the fast data reduction we have to consider other requirements like the resolution of the optics, the lightning, the necessary oversampling rate, film transport accuracy and the structuring of the symbols on film. We experienced that the approach of reducing the amount of data from the scanned data down to the information data already in the scanning device is only possible if also other challenging specifications of the scanning system are realized there. We analyze in addition the necessary composition of the scanning device to realize a fast reading of the bits on film.

Bits on Film scanning

Motivation: In our information society, the ability to reliably store digital data in the long-term is equivalent to the preservation of our collective memory. Long-term storage of large amounts of digital data is not only an expensive, but also a risky business. Today data migration is the most common strategy for storing and preserving digital data. The Task Force on Archiving of Digital Information (TFADI) has defined migration as "the periodic transfer of digital materials from one hardware / software configuration to another, or from one generation of computer technology to a subsequent generation. The purpose of migration is to preserve the integrity of digital objects and to retain the ability to retrieve, display and otherwise use them in the face of constantly changing technology."[1]

When copying data to a different medium, the digital information has to be made compatible with the new hardware and software environment; which necessitates a change in the structure of the original data. Herein lays one of the major disadvantages of migration: the data could become damaged, intentionally or accidentally manipulated or even completely lost. The preservation and authenticity of the data are therefore at risk during each migration cycle. Considering the exponential increase in the amount of data this strategy quickly becomes extremely costly, the more so as each migration to a new medium involves both the newly produced data as well as the old, non-copied data from previous migrations.

One interesting way of realizing long term storage of digital data is the recording on long term stable film material which is a real WORM medium and which has almost no degeneration over periods of several hundred years. The digital data can be stored there as small dots with different grey tone representing the bits ("bits on film"). In the past years Fraunhofer IPM has done a lot of development in this area, mainly focusing on the development of the recording systems [2] but also engaging in selected application scenarios like long term storage of digital audio data [3]. There are some systems exist which can record those "bits on film" reasonably fast. Two companies are already either in business (bitsave AG [4]) or at the start (Cinevation with its Archivator [5]). Also a lot of work was invested in the last few years in the fast and reliable data interpretation and recreation back to use in the original digital environment [6-8]. Despite of these activities, which indicate that "bits on film" can be used and is already in commercial use, there is still a lack in a scanning device able to convert the "bits on film" fast back to digital data. The economic success on large scale of the "bits on film" long term storage solution is strongly dependent on the availability of a fast reading (i.e. scanning and reconstruction) unit, which can verify the successful storage of the digital data as "bits on film" as part of the production of the data film. If such a reading device could verify the data storage in real-time the high volume storage of digital data as "bits on film" could become reality. We did research in the realization of such a scanning device and present the results, always keeping in mind that the system has to be used economically attractive. We are doing this work in an EU-cooperation project "MiLoS" [9] together with companies producing the optimized film medium and the other necessary hardware, including the company Cinevation which wants to get in business with that in short term. Therefore also the economic relevance of our investigations is obvious.

Problem: Comparing the activities of the diverse companies and research institutes in this field the bits are packed typically in small dots on the film beginning at diameters of 4µm and ending at about 10µm in order to get high data volume on the film medium. The recording system we refer to in our investigations is the "Archivator" produced by Cinevation. This system records the bits extremely fast in squares of 6µm size on 35mm roll film. An area of about 24 x 15mm² is recorded there in 40ms. That corresponds to a data recording rate of about 26MByte/s using only 1bit per pixel or corresponding 105MByte/s corresponding to 4bit per pixel (using grey level). This allows archiving huge amounts of digital data. In order to guarantee the complete storage of data you have to scan the recorded "bits on film" and compare them with the original .As actually the data rate for scanning and especially for interpretation of the data is significantly lower nowadays the data recovery is the bottleneck of the whole archival production chain. In order to realize an economically attractive archival chain it is therefore important to realize a high speed scanner which is capable of scanning and interpreting data rates of more than 40MByte/s. In figure 1 the time consumption for each step is shown. The base for these numbers is a scenario set up in the MiLoS project using the "Archivator" as recording system and scanning and reading systems which are available on the market. Even if other components will be faster e.g. in the unboxing the huge difference between recording and reading speed is obvious.

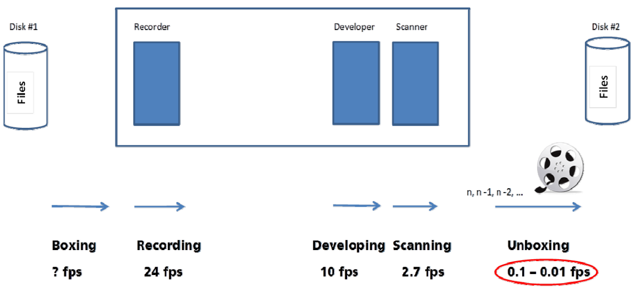

Figure 1: Principal set up of the production chain from digital data to "bits on film" and back (for verification). Even if the exact numbers can differ between various setups the huge difference between recording and reading speed is obvious.

Approach: The challenge is to build up a concept which is capable of handling this high data rates. One additional tightening aspect is next to the high data rate the ambition to increase the amount of data respectively the data density on the medium. As the data density increases quadratically with the pixel size the aim is to decrease the pixel size.

In our case in the MiLoS project the pixel size for each information pixel was reduced down to 6µm. The recording of data is done in frames with the same size of 25 x 15 mm². In figure 2 the construction of a part of the frame pattern is shown: at the edge of each frame an orientation pattern, the pixelmap, is used. Inside of that pixelmap the data is recorded as 6µm pixel with distinct grey levels. Basically to read small dots you have to use a scanning device which has a theoretical oversampling of more than two (Nyquist–Shannon sampling theorem). In order to get a robust, error free reading device you have to realize an oversampling rate which is significantly higher. Taking that into account we are using an oversampling rate of 4 in the scanning device. Therefore the amount of scanned pixels is of a factor 16 higher than the amount of original "information" pixels. Additionally to that the film is scanned with a bit depth of 8bit. For high speed scanning it is essential to reduce this amount of data significantly in an early stage. Due to that the data interpretation has to be about 30 times faster than the recording speed if you want to get back the data with the same speed as recorded.

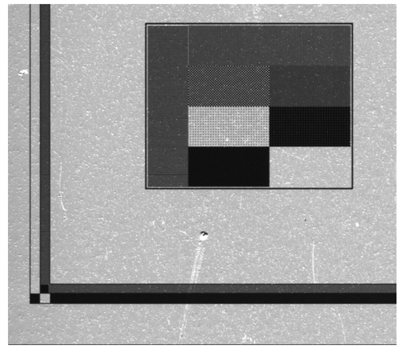

Figure 2: Picture of a part of the frame pattern. The data is recorded inside a "pixelmap" called orientation pattern which is positioned at the edge around each frame.

<u>Data reduction and recovery</u>: In order to proof the functionality of our approach we build up a test scanning device called TestScan.

Figure 3 shows the principle set up of the TestScan together with the interconnection of the hardware used and points out our new approach for the data recovery algorithm: Whereas the typical way of data handling in a scanning device is the transportation of all the raw data from the camera to the computer memory, where the data reduction is done with various algorithms we are using visual applets already in the frame grabber which is directly on the backplane of the line camera.

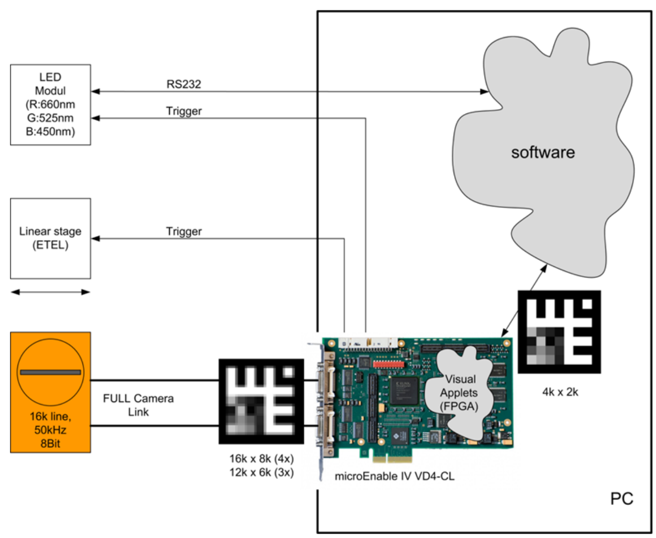

Figure 3: Block diagram of the TestScan. The high amount of data is decreased already in the "microEnable" unit with the use of data reduction software directly written in the FPGA.

The “microEnable” unit consists of two FPGA´s. The additional FPGA is used in our high speed scanning device for reducing the data line for line in real time. The actual camera system allows a scanning frequency of 50 kHz which can be increased up to 100 kHz soon. We use Visual Applets as the tool for programming the additional FPGA on the frame grabber.

The Visual Applets has a constraint which has to be kept in mind. In this context the limited capacity of the FPGA has to be mentioned. Keeping in mind that already one line of the “bits on film” consists of 4000 pixels which are oversampled with the factor 4 with a bit depth of 8bit you get even for only one line an amount of scanned data of 512kbit. This number is still significantly under the maximum capacity of the Visual Applets described above, but you always have to bear in mind that the high speed scanning system must be economically. Therefore the tolerances of the main components like mechanical film transport could be in a range of more than 10µm corresponding to the effect that the resulting tilt of the film can be over several “bits on film” lines. This requires the calculation of all these lines in the visual applets simultaneously which could lead above capacity of the Visual Applets.

In order to enable nevertheless the reliable and fast data reduction a better condition monitoring of the general requirements is of crucial help. In our experiments several assisting elements are introduced which guarantee that some important basic conditions for the scanning system are kept within the tolerances:

Pixelmap on the film: First the scanner has to get exact information about the alignment of the film in real time. This can change quite fast due to mechanical tolerances of the scanner or the film material. In figure 4 you see the alignment symbols at each corner of each frame. These points of both corners at the beginning of each frame enable the scanning system to calculate the exact alignment of the frame. This information is calculated by the FPGA in real time.

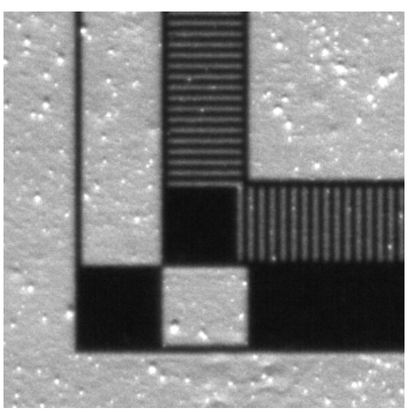

Figure 4: Each corner of the frame has these quadratic position symbols marking the “coordination system” of the frame. The point where the two squares contact each other marks the exact position. Using these points of both corners at the beginning of the frame enable the system to calculate the exact alignment of the frame.

Next to that the scanner has to know where exactly the lines on the film are positioned over the whole length and width of each frame. This can be prepared by recording a line structure on each frame of the film. Figure 5 shows this crosswalk structure which is recorded on the edges of each frame. These structures consist of alternating lines of dark and bright lines which are in line with the “bits on film” information pixels. As these structures are exactly detected by the scanning system the small misalignment of the film due to tilt, different moving speed etc. can be detected fast.

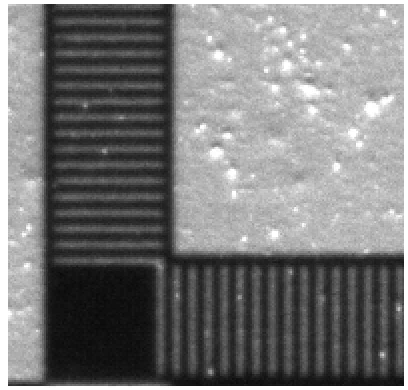

Figure 5: Crosswalk structure recorded at the edge of each frame to get fast and exact information on the alignment of the film during the scan.

These structures are used to organize the correct pixel detection and to make sure that the scanning process will be robust even if (e.g. during lifetime of the scanning unit) some distortions or misalignments occur. This information can be sent to the main control unit of the scanning system in real time, so that automatic readjustments of the system are possible.

FPGA data reduction: The pixelmap allows measuring exactly the tilt of each frame. In order to effetely use the FPGA data reduction capacity the scanned lines have to be tilted virtually to the orientation of the data lines. Ensuring that leads to the crucial step of the data reduction of the factor 32 necessary to generate the output of pure information data: due to the geometrical oversampling of the factor 16 and the scanned bit depth of 8bit using only 4bit of data level. This reduction of the 256 grey levels to 16 grey domains is done using look-up tables.

For clarification it has to be mentioned that the data interpretation (e.g. using error correction codes or the conversion back to the data file) is not part of this work. The task here is to get the “raw” information data scanned and read in real time.

Next to this crucial pixelmapping on the film also the optical and mechanical main components have to be kept in mind. In realizing the TestScan system the real time data reduction described above will result in a high speed scanning with a film transport speed of about 150mm/s for the described 6µm pixel structure!

In order to get this high speed scanning device running robust the following components have to be investigated in connection with their capability for a positive interface to the real time data reduction:

Line scanner: the number of the line size has to be 16k, as the “Archivator” system records 4k pixels per line and as we realize an oversampling rate of 4. In the TestScan we use the camera from Rauscher, ELiiXA+ C4M CL 1605 with full Camera link, 50 kHz, which can be extended to Rauscher, ELiiXA+ CXP 1605 with CoaxPress, 100 kHz.

Lightning: the film strip has to be lighted with homogenous, high intensity lightning in order to receive good images. We use a LED light converted by fibers to an effective area of 50 x 3.2 mm^2 followed by a diffusor.

Film transport: the film has to be moved precisely and continuously over the scanning line. Typical numbers of high precise transportation systems realize an accuracy of a few percent. Without correction of these tolerances the misalignment over one frame consisting of 2160 lines could be several lines. Nevertheless the pixelmap on the film described above corrects this tolerance in real time and therefore the data reduction can manage that transport tolerance.

Next to that and also quite important is to measure the contrast of the pixelmap (crosswalks) in real time in order to indicate the actual position of the film in z-axes. This is an crucial value for the correct data reduction as a low contrast handicaps the correct classification of the grey level to the grey domain. The value of the actual contrast value can be used for the main control unit of the scanner system to control the in-focal position of the film.

Scanner optics: the required MTF for the scanner optics at 80lp/mm (this corresponds to 6µm pixel size) has to be above the value of 0.8.

For a successful approach you have to consider the dependency of the above mentioned components from each other. Therefore the best composition has to be developed. Several different test structures therefore have to be scanned and the robustness of the data recovery at a fixed speed has to be optimized.

Results: The introduced new approach is still in the stage of investigation. The intermediate results will be presented and the interdependencies of the different factors and components will be explained. The chosen components will be introduced and the benefit for the high speed scanning system will be illustrated. Some scanning samples will be shown. The effect of the rotation of the image on the reduction of data will be shown.

Conclusions: This abstract outlines the new approach of the high speed scanning device which has the goal to rescan data from film with data rates of about 40Mbytes/s. As the investigations are still in progress final results cannot be presented yet, nevertheless the explanations should have shown that the method looks promising. First results show that the attempt is promising. The data reduction in real time outlines the path to enable and realize a high speed scanning system which can scan the "bits on film" with a speed for film transportation up to 150mm/s using of-the-shelf components. Further steps for a possible integration in a complete production system for writing digital data on long term stable film material for the archiving of the data will be shown in the future.

References

The project is funded by EU in Eurostar program. Project name is "MiLoS - Develop an Open Standard Migration Free Storage and Archival Medium for long term archiving of digital data". [9]

References

[1] TFADI (1996): Preserving Digital Information – Final Report. Task Force on Archiving of Digital Information (TFADI), http://www.rlg.org/ArchTF

[2] D.M. Giel, A. Hofmann, W. Salzmann, and C. Voges, „Digital Data Storage on Microfilm – The MILLENIUM Project: Hardware Realisation", in Proc. Of IS&T Archiving Conference, Arlington, VA, May 2009, pp 80 – 81

[3] A. Hofmann and D. M. Giel, DANOK: Long Term Migration Free Storage of Digital Audio Data on Microfilm", Proc. Of IS&T Archiving Conference, Bern, Switzerland, June 2008, pp. 207 -211

[4] http://www.bitsave.ch/

[5] http://www. cinevation.net

[6] Müller, F.; Fornaro, P.; Rosenthaler,L.; Gschwind, R.: PEVIAR: Digital Originals. ACM Journal on Computing and Cultural Heritage, Vol. 3, No. 1, Article 2, June 2010, pp. 2:1–2:12.

[7] Voges, C.; Märgner, V.; Fingscheidt, T.: Digital Data Storage on Microfilm – The Millenium Project: Signal and Information Processing. IS&T Archiving Conference, Arlington, VA, USA, Mai 2009, pp. 187–191.

[8] Voges, C.; Märgner, V.; Fingscheidt, T.: Digital Data Storage on Microfilm – Error Correction and Storage Capacity Issues. IS&T Archiving Conference, Bern, Switzerland, June 2008, pp. 212–215.

[9] https://www.eurostars-eureka.eu/

Author Biography

Ulrich Klocke, received his degree of Dipl. Math. in 1975 at Albert-Ludwigs-University Freiburg. He works at Fraunhofer IPM in the area of laser recording and robust spectroscopy systems.

Joachim Anders, received his degree of Physical Engineering in 1987 at Hochschule Heilbronn. Since 1990 he works at Fraunhofer IPM in the area of laser recording and robust spectroscopy systems.

Daniel Steiger, received his degree of Dipl. Engineer for Electrical Engineering in 2005 at BTU Cottbus. Since 2005 he works at Fraunhofer IPM in the area of laser recording and fast data transmission.

Andreas Hofmann, received the degree of Dipl. Phys. in 1995 at Albert-Ludwigs-University Freiburg. Since 2000 he works at Fraunhofer IPM in the area of laser recording and data storage.