Bear this in mind when selecting a book scanner

"See" what is important

Books, magazines and historical documents come in hugely different colors, shapes and sizes; for libraries, archives and museums, digitization of these stocks is proving to be a greater challenge than ever before. When setting up and investing in a suitable digitization platform, the type of scanner system to be used is an instrumental part of the selection process. The purpose of this White Paper is to provide you with basic information, which should help you to understand and pin down the key functionalities and quality features of the digitization systems and prevent you from making a bad investment.

The market basically falls into two categories of book scanning systems depending on the method used: The overhead scanner comprising a combination of line scan camera and light unit and the overhead camera as a scanner. In the case of overhead scanner, line, optical system and illumination are intrinsically linked, moving over the document as a single unit. In the overhead camera, a digital camera is centrally mounted over the document to capture the image. Illumination is provided by a light source attached to the device, some devices are supplied without their own light.

Both systems almost exclusively use CCD (Charged-Coupled-Device) sensors. CCD sensors are sophisticated technological devices, their functionality has a proven track record in a variety of different image capture systems. The main advantage of this sensor technology is its extreme high sensitivity to light, whereby disruptive interference, such as image noise, is kept to a minimum. Its structure is relatively simple, resulting in very few defective pixels in semi-conductor manufacture, whereby, owing to the lower overall number of pixels, line scan sensors are the only ones able to be delivered error-free.

CCD line versus CCD array

The difference between the two scanning systems lies in the configuration of CCD chips as a line or an area sensor. Sometimes, the latter is also known as an array sensor. In the CCD line example, the individual light-sensitive cells are placed in a row. A two-dimensional image is created by the movement of the line across the document. This movement is designed in such a way that the sensor is advanced by exactly the distance of one pixel during the integration time. In order to produce color images, pixels are typically arranged in extremely close proximity in three parallel rows. Each row has a color filter - usually red, green and blue. This gives rise to a color separation image.

The line length in standard overhead scanners is 7,500 to 10,680 pixels; in some instances, it can even exceed 20,000 pixels. In the case of a tri-linear color line with, for example, a line length of 7,500 pixels, a line scan sensor has more than 22,500 discrete light-sensitive elements. Multiply this value by the sample rate x scan path to obtain the total resolution of the system. An important development in this connection: For the resolution proposed, overhead scanners - in contrast to overhead cameras - do not use interpolation methods to produce image content.

In the case of area sensors or array sensors, individual cells are arranged two-dimensionally. To obtain color information, the cells are alternatively provided with R/G/B filters. In order to capture the full color information of an individual pixel, either the whole array needs to be shifted and read several times or the missing colors must be interpolated from information contained in the neighboring cells. The largest sensors on the market have up to 7,000 x 10,000 pixels, i.e. 70 megapixels, and are correspondingly expensive.

Expert advice: Volker Jansen, Technical Director, Zeutschel GmbH

The quality of digitized images depends on several different factors. From these, actual or real resolution, color reproduction and homogeneous illumination are the most relevant.

Accurate in every detail or roughly approximate?

In evaluating the quality of a digitization system, image resolution is a key aspect. Resolution is the ability of a system to reproduce finer or just coarser structures. The higher the resolution, the finer details or structures can be transferred from the original into the digital reproduction.

In comparing the two scanning systems, it is important to have the actual resolution, i.e. non-interpolated. In general terms, the sensor's megapixel data is used to describe the scanning system's overall resolution. In actual fact, the real resolution is significantly below this. A 40 megapixel chip contains 20 megapixels of green information plus 10 megapixels of red and blue information in each case. This means, only between one third and one quarter of the sensor's megapixel resolution are left for the actual, real resolution of the system, the remainder is interpolated. Hence, a 40 megapixel area sensor scanning system corresponds to a 10 or 13 megapixel line scan sensor scanning system. In comparison to the area sensor, the line scan sensor delivers an actual resolution that is several times higher.

Resolution: Advantage line scan sensor

An A2 size document is digitized using the two scanning systems. For the book scanner with overhead camera, a 40 megapixel area sensor conventionally used in this performance class is utilized. The color mosaic filter dispenses with 20 megapixels of green information and 10 megapixels of red and blue information respectively from the 40 megapixels. If the A2 document is scanned using a line scanner with a 7,500 pixel line length, the volume of information obtained is 10,140 x 7,500 pixels or 75 megapixels per color channel or a total of 228 megapixels for red, green and blue. So, the line scan sensor provides six times the resolution in comparison to the area sensor.
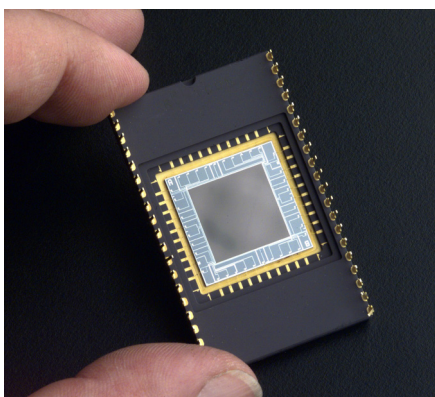
Area sensor with 40 megapixels.
Image of an A2 document = 40 MP
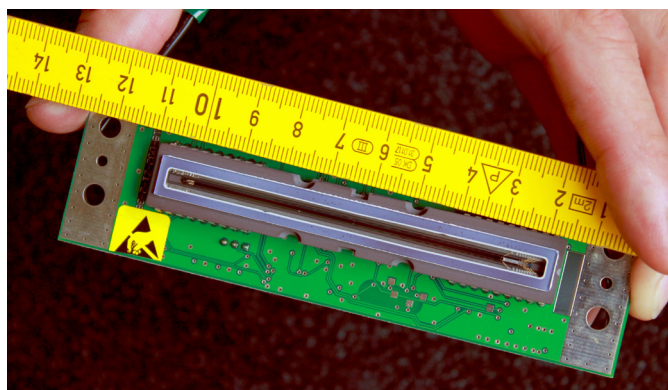
Line scanner with 7,500 pixel line length
Scan of an A2 document = 228 MP

Owing to the large number of total pixels, affordable array sensors always have pixel errors. According to the specifications of commercially available 50 megapixel area sensors, they may, for example, contain 4,000 faulty pixels, a maximum of 50 faulty clusters (collection of neighboring defective pixels) and up to 20 faulty columns. These damaged areas must be corrected in the image at a later stage using mathematical interpolation. Information from the neighboring pixels is transferred at the same time, which, in reality, is not a real correction but a rough approximation at best.

In addition, it is also a good idea to take the problem of 'color interpolation' into account. The majority of area sensor systems limit pixel by pixel sampling to one color per pixel. The missing information must be generated by interpolating the neighboring pixels. In documents with fine structures and high contrasts, this so-called 'color interpolation' results in image distortions in the form of 'moiré effects'. In its digitization guidelines, the German Research Foundation took up this problem and, therefore, includes the line scanner as an appropriate system for such things as delicate motifs and gravure printing.

Expert advice: Rüdiger Klepsch, Managing Director/ COO Marketing & Sales, Image Access GmbH

Users should not be fooled by the high megapixel information of the scanner vendors with area chips. As with digital cameras, the rule is: What counts is the actual resolution, not the interpolated one.

Realistic or just colorful?

Good or bad, sharp or blurred are judgments based on personal impressions. However, this impression varies from person to person, so this cannot be taken into account during the evaluation process. The parameters must, therefore, be measurable. These parameters also include color reproduction.

Color reproduction is a parameter to determine how exactly a system is able to reproduce a particular color correctly. Color reproduction also predicts the extent a system is able to capture colors.

The ICC standard is key in evaluating color reproduction. Manufacturers of graphics, image processing and layout programs founded the International Color Consortium (ICC)

in 1993 with a view to standardizing color management systems. So-called ICC profiles characterize the color space of the color input- or color reproduction devices. The aim is for a document captured using a scanner to be reproduced as closely as possible on a monitor or printer. Users should, therefore, take strict care to ensure the scanner used comes with an ICC profile and that the capture software supports the ICC specifications throughout.

Commercial book scanners have real color processing with an output of 24-bit color and 8-bit greyscale. High-quality line scanners scan the page to be captured with a high bit rate (for example, 42-bit color). Necessary corrections for linearity, color, homogeneity, etc. are conducted at this high bit depth. Only then is the fully corrected image reduced to the 8-bit per color channel (24-bit) output format used.

More light!

Lighting has a special role to play for the best image results. It must illuminate the document in a homogeneous way and provide enough light to suppress the extraneous light in open systems. Here, an effective way has been to increase illumination by 30 times the level of extraneous light. In addition, it must also be safe enough in conservation terms so as not to damage the document. Finally, in order to maintain a consistent level of quality, it must keep working at constant temperature levels for long periods.

Expert advice: Markus Schnitzlein, Managing Director of Chromasens GmbH, a specialist in imaging systems

Every brilliant image requires two things: An outstanding camera and outstanding illumination. Obtaining good scanning results demands sufficient and homogeneous light because high-quality images can only be captured when bright, homogeneous and stable illumination is available.

Line scanners are increasingly focusing on LED technology. LEDs command a high degree of light efficiency, good color reproduction, high spectral stability and allow linear focusing. Homogeneous illumination is made possible by arranging LEDs in rows.

This means that the influence of ambient light becomes of secondary importance. The document also benefits from only being exposed to an extremely limited amount of light. Line-based systems work with 'moving lights', i.e. a point on the document is only directly illuminated during the scanning process. High-end scanners with a luminance of up to 40,000, therefore, only illuminate the individual point for 0.2 seconds.

None of the area sensor systems currently available on the market use a comparable light, making it very difficult to attain optimal and constant levels of quality. This requires spectrally-correct continuous illumination, which, from a conservation point of view, is a cause for concern. Even when the intensity of light is low, for example, at 1,500 Lux, the document's exposure level is 25 Lux hours (Lxh), which is higher than high-end line scanners by a factor of 10. Where the effect of ambient light is strong, a reproducible quality in color and homogeneity can hardly be achieved using area sensors. For this reason, book scanners with area sensor technology have only very limited use in the open-access sections of libraries.

In summary

1) Scanner systems with CCD line scan sensors support a much higher actual resolution than currently available systems with CCD area sensors. When processing an A2 document, for example, line scan sensors with a line length of 7,500 pixels offer a six times higher resolution than 14 megapixel area sensors.

2) Although interpolation superficially increases the numerical value of the resolution, this does not apply to the quality of the raw data.

3) 'Color interpolation' often leads to image interference in the form of a 'color moiré".

4) Color management according to ICC standards is essential. In terms of software, availability of ICC profiles and consistent support of ICC specifications are obligatory.

5) Line scanners with LED lighting systems guarantee high levels of light efficiency and homogeneous illumination. Consequently, they are very good for use in a wide variety of different light conditions.

6) Scanning systems with area sensors require optimal lighting conditions with low ambient light and, therefore, can hardly be put to use in open-access areas.

Contact

Zeutschel GmbH, Heerweg 2, 72070 Tübingen, Tel.: +49 711 / 9706-0, info@zeutschel.de, http://www.zeutschel.de

Image Access GmbH, Hatzfelder Strasse 161-163, 42281 Wuppertal, Tel: +49 202 270 58-0 info@imageaccess.de, http://www.imageaccess.de