



To bear in mind when selecting a book scanner

## „See“ what is important

**Books, magazines and historical documents come in hugely different colours, shapes and sizes. For libraries, archives and museums, digitisation of these items is proving to be a greater challenge than ever before. When setting up and investing in a suitable digitisation platform, the type of scanner system to be used is a fundamental part of the selection process. The purpose of this White Paper is to provide you with the basic information to help you to understand and pin down the key functionalities and quality features of the different digitisation 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 the 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 CMOS-sensors (Complementary Metal Oxide Semiconductors) and 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 both sensor technologies are their very high sensitivity to light, whereby disruptive interference, such as image noise, is kept to a minimum. A CCDs 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 – CMOS and CCD - are the only ones able to be delivered free of defect pixels.



*Zeutschel Overhead scanner OS 16000 Advanced Plus*

#### **CCD line versus CCD array**

The difference between the two scanning systems lies in the configuration of the single pixels on the chips as either a line or an area sensor. Sometimes, the latter is also known as an array sensor. In the CCD line sensor 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 colour images, pixels are typically arranged in extremely close proximity in three parallel rows. Each row has a colour filter - usually red, green and blue. This gives rise to a three-colour channel RGB image. The line length in standard overhead scanners varies from 7,500 to more than 16,000 pixels; in some instances, it can even exceed 32,000 pixels. In the case of a tri-linear colour line sensor 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 sampling rate (DPI) × scan length (Inches) to obtain the total resolution of the system. Important in that case is that for the resolution proposed, overhead scanners - in contrast to overhead cameras - do not use colour interpolation methods to produce RGB information, since for any single pixel Red Green and Blue is detected.

In the case of area sensors or array sensors, individual cells are arranged two-dimensionally. To obtain colour information, the cells are alternatively provided with R/G/B filters. In order to capture the full colour information of an individual pixel, either the whole array needs to be shifted and read several times or the missing colours must be interpolated from information contained in the neighbouring cells.

The largest sensors on the market have up to 10,000 × 15,000 pixels, i.e., 150 megapixels, and are correspondingly expensive.

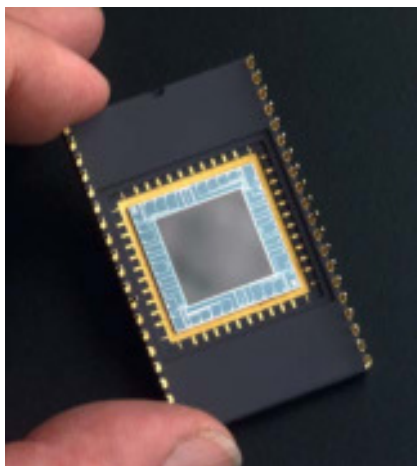
**Accurate in every detail or roughly approximate?**

In evaluating the quality of a digitisation 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.

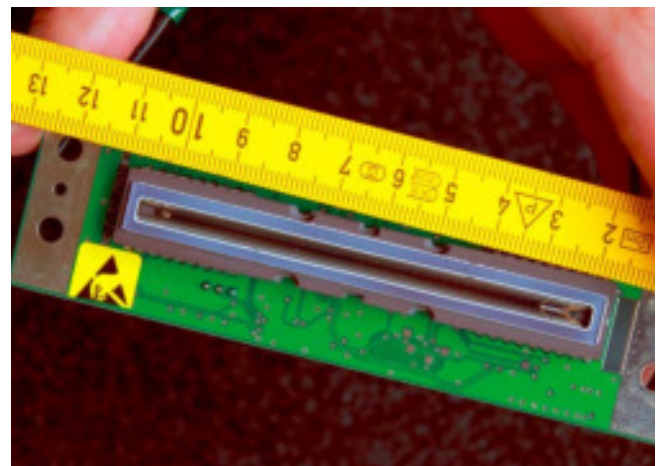
In comparing the two scanning systems, it is important to have the actual resolution, i.e. non-interpolated. In general terms, the sensor's megapixel data is used to describe the scanning system's overall resolution. In actual fact, the real resolution is significantly below this. A 40-megapixel chip contains 20 megapixels of green information plus

10 megapixels of red and 10 megapixels of blue information. This means that 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 delivers approximately only a third of its megapixel in resolution without information loss caused by colour interpolation.

Since a line scan sensor scanning system detects RGB on each pixel it delivers a higher resolution in an image with the same total number of pixels produced with a colour array sensor.



*Area sensor with 40 megapixels.  
Image of an A2 document = 40 MP – 20MP  
for Green, 10MP each for Red and Blue*



*Line scanner with 7,500 pixel line length  
Scan of an A2 document = 75MP information for each Red, Green and Blue. Corresponding to an Array sensor image with 225MP*

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

### Realistic or just colourful?

“Good or bad” and “sharp or blurred” are judgements 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 colour reproduction.

Colour reproduction is a parameter to determine how exactly a system is able to reproduce a particular colour correctly. Colour reproduction also predicts the extent a system is able to capture colours.

The ICC standard is key in evaluating colour reproduction. Manufacturers of graphics, image processing and layout programmes founded the International Color Consortium (ICC) in 1993 with a view to standardising colour management systems. So-called ICC profiles characterise the

colour space of the colour input- or colour 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 colour processing with an output of 24-bit colour and 8-bit greyscale. High-quality line scanners scan the page to be captured with a high bit rate (for example up to 96-bit colour). Necessary corrections for linearity, colour, homogeneity, etc. are computed at this high bit depth. Only then is the fully corrected image reduced to the 24 or 48 Bit RGB output format used.



### Expert advice:

**Volker Jansen, Technical Director, Zeutschel 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. This can be measured along international accepted methods described in standards like e.g. ISO 19264-1

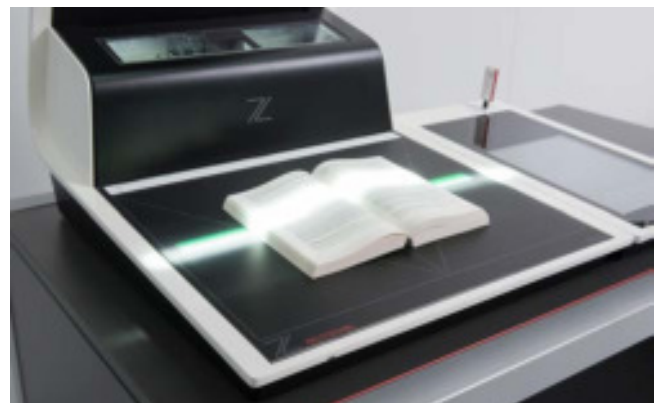
### More light!

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.

Line scanners are increasingly focusing on LED technology. LEDs command a high degree of light efficiency, good colour 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 being scanned 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.

The most important fact is that the light has to be ergonomic and safe for the operator. This can be verified by testing the light system against standards like IEC EN 62471 “Photobiological safety of lamps and lamp systems”.



*Zeutschel zeta*





*Zeutschel ScanStudio*

Only one of the area sensor systems currently available on the market use a comparable lighting technology (Zeutschel ScanStudio) to obtain optimal and constant levels of quality.

An array sensor system 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 colour 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.

#### To sum up

- 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 significantly higher resolution than 40-megapixel area sensors.
- Although interpolation superficially increases the numerical value of the resolution, this does not apply to the quality of the raw data.
- 'Colour interpolation' often leads to image interference in the form of a "colour moiré".
- Colour management according to ICC standards is essential. In terms of software, availability of ICC profiles and consistent support of ICC specifications are obligatory.
- 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.
- 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.