How can image scanners read optical forms?

Image scanners are devices that scan papers and produce digital images. Image scanners classified as "document scanner" are more suitable for optical form reading because these devices involve an automatic document feeder.

Document scanners are home and office devices that serve multiple purposes:

- Digital archiving of documents
- Information capture
- Document transmitting and sharing
- Copying

Document scanners can also be used for optical mark reading which is a type of information capture.

Traditionally optical mark reading is realized using a dedicated hardware called Optical Mark Reader, which is designed to read optical forms accurately and fast. Today, image scanners provide much better speeds than their predecessors. Therefore detecting marks on the scanned image is now feasible instead of installing special hardware for detection.

Mark sensing from scanned images is a software method. Tools used for such purpose are called "OMR Software".

Basic functions of OMR Software

The basic function of OMR software is to:

- capture images from scanner
- process images using image processing techniques
- locate mark positions and read mark states
- convert mark states into final data.
- store, export, convert final data for application software.

This way an image scanner can replace an OMR scanner. OMR software can go beyond data capture and present some analysis features too (MarkReader can score exams, provide item analysis and provide reports for surveys). Because OMR software is based on detailed image of forms rather than mark states, OMR software can also present other capture capabilities. OCR and ICR can also be used besides OMR, barcode and QRCode can be detected, image clips (such as photos, signatures) can be extracted. This capabilities has a number of advantages over OMR when identifying forms and collecting more sophisticated information.

Difficulties

There are difficulties encountered during reading of forms from scanned images:

- Scanners might not feed papers correctly. Rotated and tilted images have to be corrected to align mark positions with the form template. MarkReader uses timing marks at both sides to align image

with the form template. Scaled, rotated, tilted images are precisely matched to form design.

- Darkness/brightness of scans vary with respect to scanner and scanning options which may lead to incorrect reading. MarkReader uses drop out colors technique to remove any information other than marks which reduces risk of false reading.
- Erasures, dust and image noise may cause false positives. MarkReader does not rely on scanner to get the binary (black and white) image, instead full intensity is used to sense mark states. Multiple thresholds are applied to detect existence of marks. Checks and crosses are also detected and (optionally) considered as a positive mark.

Image scanner specifications

There are many scanner alternatives that vary according to the following specifications:

ADF	Automatic document feeder is a device part that feeds documents to read-head from an input tray. OMR readers are required to process multiple forms automatically, so flatbed scanners are not suitable for that purpose. Capacity of ADF trays is also an important specification when high volume scanning is needed.
DPI	DPI (Dots per inch) represents resolution of produced images. As DPI gets higher, the images captured have more details, and more size. Accuracy of OMR reading will be better if a higher dpi is selected, but scanning speed may decrease. Scanners generally specify speed according to dpi value.
Format	Format of captured images, such as JPEG, TIFF, PNG, BMP, PDF. Some formats allow multiple pages in a single file.
Color depth	There are scanners that can scan black-and-white or gray-scale images only. Some can capture color images of papers. Some OMR software prefer black-and-white scans. MarkReader prefers color scans for better accuracy, although gray scale forms and gray scale scans can also be processed.
PPM	PPM (Paper per minute) is the speed of scanner. Scanner speeds range from 2 PPM to 300 PPM roughly. Scanners that can scan both sides (duplex scan) also specify IPM (images per minute), which is generally 2 times the PPM value, because 2 heads scan the page in parallel.
Interface	Computer-scanner communication is realized using a number of industry standards. TWAIN, WIA, ICA, ISIS, SANE are among these standards.

Last modified on 2020-08-18 by admin