ICR, OCR, and OMR

A Comparison of Technologies

OMR, OCR, and ICR technologies all provide a means of data collection from paper forms. All three technologies may require imaging scanners. OMR may also be done using an OMR (discrete read head) scanner such as the OpScan® or EZData™ product lines.

Optical Mark Recognition (OMR)

OMR technology detects the absence or presence of a mark, but not the shape of the mark. Pearson NCS software interprets the output from the scanner and translates it into the desired ASCII output. Forms are scanned through an OMR scanner. The forms contain small ovals, referred to as 'bubbles,' that are filled in by the respondent. OMR cannot recognize hand-printed or machine-printed characters.

An OMR scanner can maintain a throughput of 1,800 to 10,000 forms per hour. This activity can be controlled and processed by a single workstation, which can handle any volume the scanner can generate. Increasing the throughput simply requires upgrading the scanner. ICR and OCR software cannot be used with an OMR scanner.

OMR is the fastest and most accurate of the data collection technologies. For test answers and short grids like ID number, it is also easy for the user. The accuracy of OMR is a result of precise measurement of the darkness of a mark, and the sophisticated mark discrimination algorithms for determining whether a mark is an erasure or a mark.

Intelligent Character Recognition (ICR)

ICR technology gives scanning and imaging systems the ability to turn images of hand-printed characters (not cursive) into machine-readable characters. Images of hand-printed characters are extracted from a bitmap of the scanned image. Forms can be scanned through an imaging scanner, faxed, or computer generated to produce the bitmap.

ICR is less accurate than OMR and will require editing to achieve perfect data. However, when compared to manual key entry, ICR does reduce keystrokes for data collection and provides a more friendly form for demographic data.

Optical Character Recognition (OCR)

OCR technology gives scanning and imaging systems the ability to turn images of machine-printed characters into machine-readable characters. Images of machine-printed characters are extracted from a bitmap of the scanned image.

OCR is less accurate than OMR but more accurate than ICR. OCR can provide a good alternative to preslugged OMR grids such as student ID number, thereby freeing up space on the form for other design elements.

Comparisons

The differences between ICR, OCR, and OMR technologies are demonstrated in the following areas: Forms, Scanners, Accuracy, and Storage.

Forms

OMR forms are very specialized documents. The timing track along one edge of the form indicates to the scanner where to read for marks and for indicating where to clip images. An OMR form also contains Form ID marks, which look like black boxes on the top or bottom of the form. Additionally, the cut of OMR forms is extremely precise and the bubbles must be located in exactly the same location on each form.

OCR/ICR data collection technology also has strict forms design requirements but is a bit more forgiving. There are no timing tracks but registration marks, such as triangles, are required on the document. Form IDÌs can be a design element of the form such as a form name. Documents that will have information extracted and sent to an ICR engine contain constrained hand-print boxes; respondents fill in one letter per box. When OCR/ICR are used real time with the scanner, the OMR forms design requirements apply.

On both types of documents, color is very important. The use of drop-out color in an image scanner reduces the size of the transmission between a scanner and a computer and increases transmission speeds. The use of drop-out color on an OMR scanner allows for the use of pens (in addition to pencils) for form completion. In addition, consistent paper of proper reflectance and fluorescence influence readability. On both the OMR and imaging scanners, the combination of drop-out colors and consistent paper enhance the accuracy of results.

Scanners & Software

OMR can be done with discrete head OMR scanners and driver software such as ScanTools® II and ScanTools Plus. OMR can also be done using an OpScan *i*NSIGHT™ or 5000*i*® scanner and Image ScanTools or ScanTools Plus. In either case, the same mark discrimination logic provides a reliable translation of the respondentÌs marks. Use of the image scanners will also allow data collection of ICR and OCR using the Real Time OCRÙ optio n for Image ScanTools™, ScanTools Plus, or NCS Accra™ software.

Storage and Retrieval

If documents must be electronically stored and maintained, imaging is required. With imaging scanners, images can be captured, indexed, and written to hard disk or other storage media. The images can then be retrieved and printed as needed. With OMR technology, images of documents are not created by scanners, so electronic storage and retrieval is not possible.

Accuracy

With a quality printed document, OMR technology can consistently provide 99.9% accuracy on read data. ICR and OCR technologies can also provide 99.5% accuracy if the system is tuned properly, the forms are well designed, the characters are written cleanly and neatly, and contextual editing is used. This is approximately the same accuracy achieved by data entry clerks.

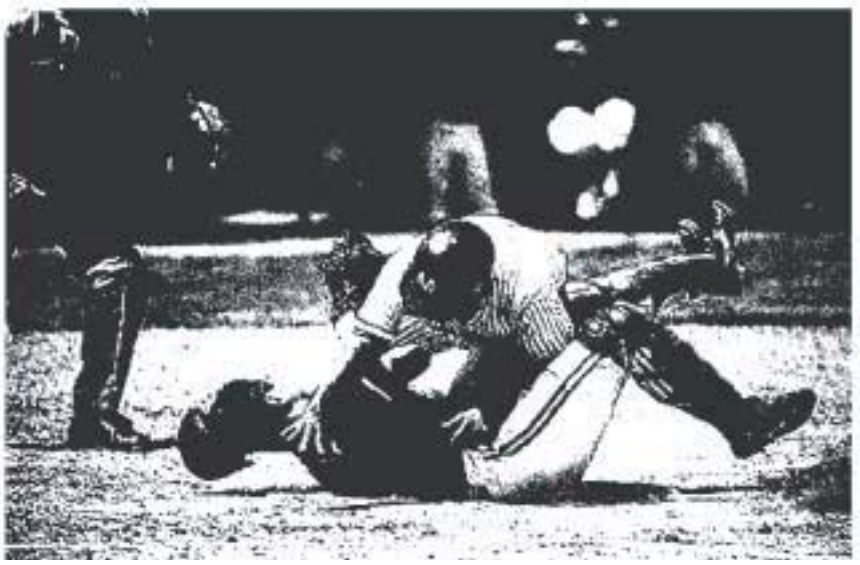
Quick Reference Chart

	Imaging	OMR
Recognition of bubble marks	Y	Y
Recognition of bar code	Y	Y
Handprint Recognition	Y- ICR	N
Machine print Recognition	Y- OCR	N
Recognition of checks and "X"s	Y	N
Requires Timing Tracks	N	Y
Requires Registration Marks	Y	N
Requires Form ID Marks	Y	Y
Accuracy	Up to 99.9% with editing	Consistently 99.9 +%
Electronic Storage and Retrieval	Y	N
Speed	1,800 Ò 10,000 sph	2,000 - 10,000/hr per scanner

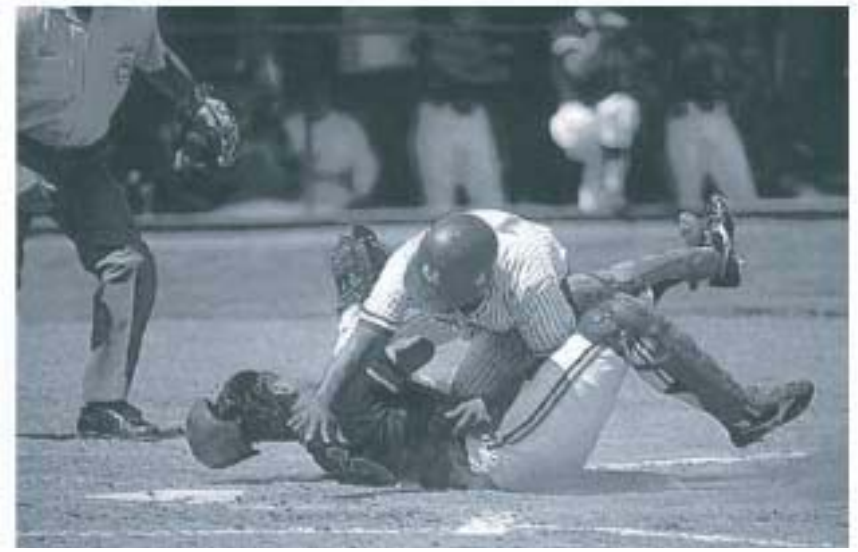
OMR vs. Mark Sense

Many image software vendors provide the ability to detect marks on a form. While it looks a lot like OMR, mark sense technology is quite different and provides less accurate results than OMR. One reason it is less accurate is that mark sense software works from a bitone image, where every dot or pixel on a page is either black or white. This results in reduced information to work from. As you can see from the following examples, there is a great deal more information available from a grayscale image.

Bitone Image



Grayscale Image



Now notice the difference in a test. The item on the left is bitone. The one on the right is grayscale.

)

The second reason that OMR is more accurate is mark discrimination. OMR sees the differences between marks and erasures. All other software simply chooses the one that has the darkest pixels. In the first example above, the erasure would actually look darker to the software.

Combined Technologies

Pearson NCS offers two scanners with software that will allow use of the best of both technologies: the 5000i scanner and the OpScan iNSIGHT scanner. For areas such as test questions responses, OMR accuracy and mark discrimination can be used. For areas such as biographical information like a student ID number, ICR or preprinted OCR may be used. Image ScanTools or ScanTools Plus software used with Real Time OCR can resolve the data and produce a single output record.

Forms Processing for ICR

For applications that require extensive ICR or OCR such as enrollment forms, Pearson NCS offers the robust forms processing software NCS Accra™. With extensive contextual editing tools, character recognition accuracy improves significantly. In addition work can be managed on one station or scale to many stations to handle peak period processing.

Storage and Retrieval of Images

Several options are available for storage and retrieval of images. The most simple is a feature of ScanTools Plus software called *i*NAME™ that allows you to intelligently name files with data from the forms so that retrieval can be done using Internet Explorer™.

For complete document management of images, ApplicationXtender™ (AX) is the answer. The software provides complete document indexing capabilities, simple or complex security schemes for users and documents, and full features image viewing capabilities. Images may be accessed either on the network or via a browser.

For more information, contact Pearson NCS at 800-447-3269, or visit **www.PearsonNCS.com** or email **info-ncs@pearson.com**.