



Data Mining

Over the years SPIE has organized more than 3000 conferences and with few exceptions each conference produced a Proceedings volume. When combined with the articles published in this journal and its sisters, SPIE publications represent a major contribution to the optical engineering literature. But the number of copies for most Proceedings runs into the hundreds, not thousands. While copies of the Proceedings and *Optical Engineering* exist in many libraries and the shelves of individual researchers, for some people, they may as well have never existed—their presence detected only by a passing reference in a contemporary paper. There are a number of good articles in there, but as time goes by, it becomes harder to dig them out.

The problems of managing the information we have and finding the most convenient way of storing it (September editorial, “Dead Tree Syndrome”) are small compared with managing the mountain of information that SPIE generates. Consider that this year’s SPIE Photonics West conference generated about 70 Proceedings, which measure 52 linear inches on a bookshelf. The new stuff is easy to find because the titles and abstracts are easy to come by. But what about the earlier publications?

Recently I was asked by SPIE to dig through their mountain of material. The goal was to select a series of papers on lens design from the SPIE literature. These papers will be scanned, formatted as Adobe Acrobat™ files, and recorded on a CD-ROM. The documents will be linked to the table of contents and will be fully searchable. This process of reclaiming such information from an archive is called “data mining.”

The first time I mined for data was during a sabbatical at the University of Arizona during the 1985–86 academic year. I had just finished *Elements of Modern Optical Design* and was painfully aware that the book did not contain anything on optomechanical design. The reason was that I didn’t know anything about it. To remedy this I decided to teach myself by reading the literature and then use my sabbatical to assemble a list of references on the subject. At that time my search resources were exceedingly limited: references from other articles, indexes

in the December issues of journals, and two private bibliography lists on similar subjects.

The contrast with my just-completed project is dramatic. The volume of material has grown rapidly since then, but so have the tools to search for the relevant documents. After doing a thorough search of SPIE publications using the InCite database on the Web, I displayed the table of contents with titles and abstracts for any pertinent Proceedings from 1990 on. On-screen searches in Netscape provided additional hits. Beyond this, one relies on a certain amount of cunning (crossword puzzles), experience (six years as a library page), and serendipity (papers that are gems located next to papers that turn out to be cut glass).

But you can only take a web search so far. Eventually I had to visit “The Yellow Submarine.” This is what I call the library at SPIE headquarters in Bellingham, Washington. Picture a modest size room with four floor-to-ceiling bookcases along facing walls and three sets of waist-high books back-to-back in the center of the room. Populate these shelves with yards and yards of yellow books. On a sunny day, the effect can be blinding. At one time some thought was given to providing a few pairs of sunglasses at the door, but because SPIE is in the Pacific Northwest, the number of sunny days is so few that it was found that they were not needed.

Because the subject was lens design I already knew where several of the “daughter lodes” were located: the International Lens Design Conferences (ILDC). (More recently the term “Lens” has been replaced by the more comprehensive one, “Optical.”) Starting with volume 237 (Mills College) in 1980, SPIE co-sponsored conferences in 1985 (Cherry Hill) and 1990 (Monterey) and published the Proceedings for each conference. The most recent volume from the recent Kona conference will also be published by SPIE. Only the 1994 Rochester conference is not part of the mining claim.

It might seem that one of the limitations on the contents of the CD would be the fact that SPIE didn’t begin to publish *Optical Engineering* (originally titled *Society of Photographic Instrumentation Engineers Journal*) until 1962 and the Proceedings in 1964. But in my digging it became clear that before 1980 there was very little that was relevant to today’s student of lens design. This was

not because the work was poorly done at the time, but rather because it was rapidly superseded by the technology of later times.

If there was a watershed event in lens design it was the ILDC held at Mills College in 1980. This was a time when personal computers were becoming widely available—when one no longer had to submit a box of punch cards or send a batch file to a large mainframe computer located in a central facility. With few exceptions most of the gold that I was mining had turned to dross at about the 1980 level. The personal computer had liberated designers from mainframes just as in the previous era the mainframes increased the range of those who had been using spreadsheets and human computers.

After all that digging I excavated a mound of ore more than 1000 citations high. The problem was then one of refinement. What of all this should be included on the CD-ROM? The answer is tied up in another question: What is this CD for? My professorial inclinations took over. To me the CD-ROM is a teaching tool. As I noted earlier, much of what is recorded in our literature is eventually superseded by more recent work. It seemed to me that the most useful papers would be those that by exposition or example provided insights into the practice of lens design. So the CD-ROM consists of 300 papers that span the topic from “aberrations” to “zoom.” Because

the papers can be organized in a number of different ways on the CD-ROM, I have added a section of tutorial papers and two sections on the most interesting and prolific contributors to the lens design literature published by SPIE over the years, David Shafer and the late Michael Kidger.

In the search through the lens design literature there were quite a few papers that described the latest improvements in a software program. Because these are superseded by a similar paper a year or so later, I did not include any of these in the collection...with a few exceptions. At the 1980 and 1985 ILDCs, there were a number of papers given that described ray tracing on some of the new microcomputers. As a look back at where we have been, I put together a little History section to celebrate some of the “early” platforms for ray tracing. Years from now, provided you can still find a CD reader for the disc, you can gesture to your lightweight 48” wall-mounted display and bring up one of these files and explain to your younger colleagues how difficult it used to be to trace rays in the “good old days.”

Donald C. O’Shea
Editor