Piecing together the priceless 'Cairo Genizah'

October 18 2011

This is a fragment from the Cairo Genizah. Credit: American Friends of Tel Aviv University (AFTAU)

A well-known collection of historical texts, the Cairo Genizah is one of the most valuable sources of primary documents for medieval historians and religious scholars. The 350,000 fragments found in the Genizah include not only religious texts, but also social and commercial documents, dating from the 9th to 19th century. But the collection is scattered among 70 institutions worldwide, including libraries in Cambridge, Jerusalem, and New York City, and scholars are hampered by both the wide dispersal of the collection as well as their fragmentary condition.

Now researchers at Tel Aviv University are working to piece together this illuminating collection, bringing the pages of the texts back together for the first time in centuries. The results are being made available to scholars around the world through a website. Profs. Lior Wolf and Nachum Dershowitz of TAU's Blavatnik School of Computer Science have developed sophisticated software, based on facial recognition technology, that can identify digitized Genizah fragments thought to be a part of the same work and make editorial "joins."

Their technology was developed in close collaboration with the Friedberg Genizah Project, a non-profit organization that seeks to facilitate Genizah research by tracking, cataloguing, and digitizing all the fragments of this collection. The research was presented at the 2011 IEEE International Conference on Computer Vision.

Restoring history

Under Jewish law, religious texts cannot simply be thrown away once they're "worn out" from overuse. While many texts were buried, many synagogues also operated genizahs, or storerooms, to store disused holy texts, usually until burial. But the Cairo Genizah, originally located in the loft of the ancient Ben Ezra Synagogue and discovered in the late 19th century, contains more than decrepit prayer books.

The genizah in Cairo became a place to dispose of texts that were not just religious in origin, explains Prof. Wolf, such as merchant's lists, divorce documents and personal letters, spanning hundreds of years. It is the largest and most diverse collection of medieval manuscripts ever discovered. For this reason, he notes, the Genizah is an invaluable resource not just for Jewish studies, but also for the socioeconomic conditions of Middle Eastern life.

In conjunction with the Friedberg Genizah Project, which has received

permission to digitize most of the fragments of the Genizah collection worldwide, Profs. Wolf and Dershowitz are working to put the pieces back together — no easy task, given the dispersal of fragments around the globe. Whereas scholars concentrate primarily on content, the software looks at features of the writing itself, since it cannot read what is written. Using computer vision and image processing tools developed at TAU, the software analyzes fragments based on parameters such as the handwriting, the physical properties of the page and the spacing between lines of writing. The program scans digitized fragments for "matches," and joins them together in a kind of digital loose-leaf binder. "Its big advantage is that it doesn't tire after examining thousands of fragments," Prof. Dershowitz says. A scholar must then review and verify the computer-proposed "joins."

So far, Prof. Wolf says, the researchers have had a great deal of success. Within a few months, they made some 1000 confirmed "joins", almost as many as were made in 100 years of Cairo Genizah scholarship. One exciting find, he notes, was the identification of pages from a work by Saadia Gaon, a prominent rabbi and philosopher from the 10th century. "All extant specimens of his work were thought to have been already discovered," he explains.

Tackling the Dead Sea Scrolls

Their work on the Cairo Genizah has extensive implications for scholars, who will have access to complete digital documents from the collection for the first time. Digital reconstructions will be publically available through the Friedberg Genizah Project Web site, a non-profit venture.

But Profs. Wolf and Dershowitz don't plan to stop with Cairo. They recently began to apply their technology to the reconstruction of the Dead Sea Scrolls in a project spearheaded by the Internet giant Google. "It's a more complicated challenge. The fragments are for the most part

much smaller, and many of the texts are very unique," explains Prof. Wolf. "These texts shed light on the beginnings of Christianity."

Provided by Tel Aviv University