The Stasi puzzle

16,250 bags of shredded documents from the former GDR's Ministry of State Security, the "Stasi", await reconstruction – a gigantic puzzle almost impossible to master by hand. Scientists at the Fraunhofer Institute for Production Systems and Design Technology IPK in Berlin have set up a pilot project to piece together this mountain of shreds in an automated process.

Even in today's age of e-mail and Internet, paper is an important carrier of secrets. This was all the more true in the former GDR. In a bid to guard the secrets of the regime, files were systematically destroyed between the fall of 1989 and January 1990 at the former Ministry of State Security. The volume of these files was so vast that the electric shredders were unable to cope, and a large part of the documents had to be torn to pieces by hand. Around 45 million A4 pages were shredded into 8 to 30 fragments each.

Only a small part of these documentis has been reconstructed so far, because it takes so long to piece them together manually: 30 people would have to work about 600 to 800 years to put together these 600 million-odd paper scraps by hand. Researchers at the IPK have found a much faster way of doing it: They have developed an automated, computer-based reconstruction process which will make it possible to analyze the documents within a short space of time. The IPK had already demonstrated the basic feasibility of such a virtual puzzling process back in 2003, in a competition organized by the Federal Commissioner for the Records of the State Security Service of the former German Democratic Republic, BStU. The pilot project for this computer-based reconstruction process is now about to begin.

"I see the automated reconstruction of Stasi documents, which has now been set in motion, as being an important contribution towards shedding light on crimes committed by the Stasi," says Klaus-Peter Willsch, member of the Bundestag, the budget committee and the Senate of the Fraunhofer-Gesellschaft, and a supporter of the project. "We must prevent investigations concerning the history of the undemocratic SED state from coming to a halt."

This opinion is shared by the committee, which has now given the go-ahead for the pilot project. Over the next two years, the contents of 400 bags will be processed at the IPK. A separate working group set up for this purpose, consisting of 25 staff members, will optimize the scanning technology employed as well as the algorithms used for image processing, and will adapt them to the volume of the paper scraps being reconstructed. "In a wide range of projects, we have developed methods of automatically recognizing and analyzing color, texture, shape, typescript and handwriting," explains Dr. Bertram Nickolay, head of the department for security systems at the IPK. "These are in addition to our techniques for automation or the learning of systems."

## Automated puzzling based on scanned paper scraps

The automated puzzle process builds on this knowledge. Before the process can begin, both sides of each fragment must be digitalized. This scanning process will be carried out by arvato direct services GmbH, who belong to Bertelsmann AG. arvato direct services and the IPK have been jointly working on the digitization of different
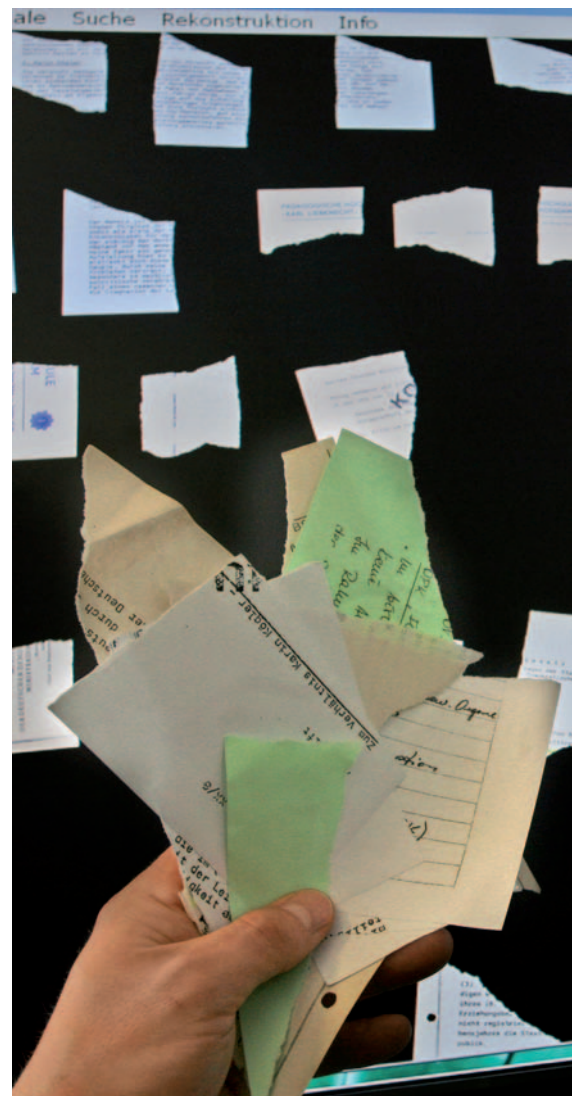
Original torn documents and their digitalized image.
© dpa/Stephanie Pilick
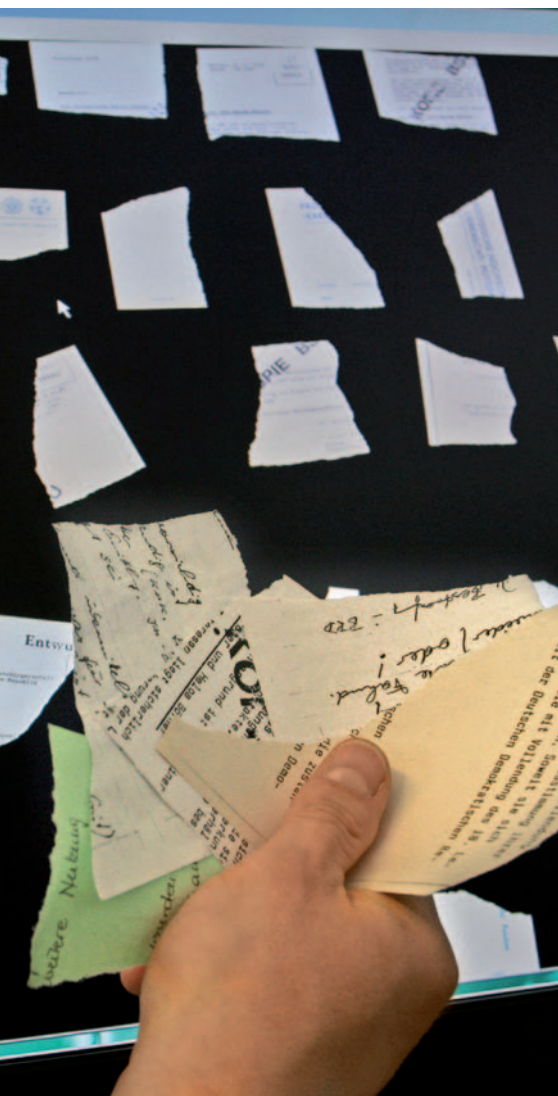
kinds of documents since 2005. This collaboration, which is supported by the state of Berlin, has produced novel scanning concepts which are now going to be used to digitize the paper fragments. The special feature of these concepts is that the pieces of paper can simply be fed into the scanner without first having to be laminated or treated in any special way. Both sides of the paper are recorded in an instant. The digital images can then be analyzed, sorted into categories and stored in a database.

Searching for shapes, colors and motifs

"Virtual puzzling follows the logic of manual puzzling," explains IPK project manager Jan Schneider. Humans use a variety of features to decide whether two pieces of a jigsaw puzzle belong together or not, such as the shape of the pieces and the colors or motifs that can be seen on each piece. This pre-selection process makes it easier to find matching pieces of the puzzle. "The virtual puzzle process begins in the same way," says Jan Schneider. "The system evaluates various describing characteristics such as shape or texture in order to reduce the search area. The actual puzzling process then takes place in this reduced area." The contours of the fragments are compared with each other and checked for matches.

If matching parts are found, they are merged to form a larger piece. Then the whole process is repeated. Piece by piece, page by page, the Stasi documents will thus be reconstructed.

Meanwhile, the researchers at the IPK are one step further than in 2003 in terms of developing the technology: Their algorithms can now reconstruct not only manually shredded documents but also those shredded by machines. This is particularly complex, given that machine-shredded paper lacks an important feature required for the puzzling process: shape. Instead, fragments of characters have to be used as the main feature. However, the researchers have been able tofully reconstruct a bag of machine-shredded documents for a tax investigation office.

The pilot project for the automated reconstruction of Stasi documents will run for two years. In parallel, historians will analyze the contents of the assembled pages. Afterwards, the decision will be made as to whether to begin with the main project in which the remaining 15,800 bags are to be puzzled together. Luckily, the members of the Stasi did not anticipate today's technological developments, which will now help to expose their machinations.
Beate Koch

Even machine-shredded papers can be reconstructed (figure above).
© dpa/Stephanie Pilick

A conveyor belt will soon transport the paper fragments to the scanner (figure on left).
© dpa/Stephanie Pilick

Both sides of each piece are scanned in one go (figure above).
© dpa/Stephanie Pilick