



**AUTOMATED VIRTUAL  
RECONSTRUCTION OF  
FRAGMENTED DOCUMENTS**

# DESTROYED CULTURAL ASSETS ARE REBORN



Fraunhofer IPK in Berlin stands for excellence in production science with a strong emphasis on digital solutions – as well as for a unique technology for virtual reconstruction of destroyed documents and artifacts. It is of vast interest for cultural as well as forensic institutions around the globe.

Paper, papyrus and parchment are highly significant cultural mediums. They carry information, they can function as valuables or store knowledge and memory – an individual person's or a collective's. Written documents fulfill communication tasks not only at the time they are created and distributed, but sometimes as much as centuries or even thousands of years later, when following generations study their ancestors' remnants for historic research purposes. But if these remnants get intentionally or unintentionally damaged, ripped or shredded, until now the information recorded on it was lost whenever restoration proved either too time-consuming or too expensive – if it was possible at all.

Starting in 2006, research at Fraunhofer IPK has fundamentally changed this: Fraunhofer IPK experts have been developing a

technology that enables computers to »puzzle« damaged documents back together from the pieces they were torn into. Originally designed to make ruptured files of the German Democratic Republic's secret service legible again, the »ePuzzler« technology has meanwhile been applied to many contexts, not only in Europe, but also overseas. It helps to reconstruct documents such as those from the Historical Archive of the City of Cologne that collapsed in 2009 and from the Asociación Mutual Israelita Argentina (AMIA) building in Buenos Aires that was destroyed in a bombing in 1994. It is also used to put together shreds of documents that serve as evidence in forensic contexts. And it is no longer limited to two-dimensional material such as paper, but can also be applied to reconstruct destroyed plane frescoes and other 2.5D cultural assets.

## Production technology applied in a cultural context

At first glance, there seems to be little connection between production technology and historical documents – so why would a research institute dedicated to production technology get involved in document reconstruction? The answer lies in Fraunhofer IPK's expertise in image analysis and computer vision. The ability to process visual information is a basic requirement for many auto-

mated processes. A machine can only work autonomously, if it can »recognize« a component, its position and condition. Therefore, image processing and pattern recognition have been areas of research at Fraunhofer IPK since the 1980s. They are applied in inspecting material surfaces for quality assurance, in controlling handling and assembly processes and in monitoring systems that ensure the safety of workers in hazardous areas.

It was by personal interest and commitment of Fraunhofer IPK researcher Dr. Bertram Nickolay that the knowledge acquired in these areas was applied to a cultural context. The Machine Vision department headed by Dr. Nickolay is now internationally acclaimed for image analysis for virtual reconstruction.

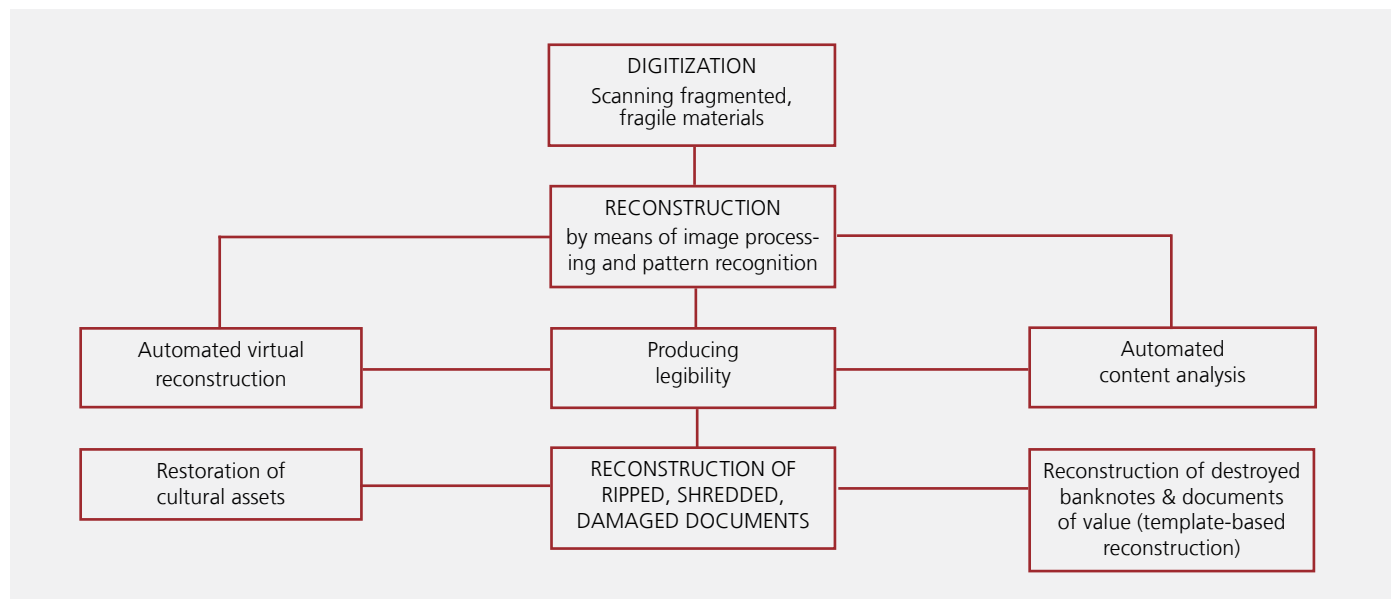
### The reconstruction process

The reconstruction process comprises of three steps: To process fragments with a computer, first they have to be fed into it – meaning they must be digitized. With the vast variety of digitization systems available on the market, this seems like a negligible aspect. However, it became evident in the very early stages of research that digitization

would become a topic that had to be specifically addressed by Fraunhofer IPK's development activities.

After digitization, ePuzzler evaluates the fragments' digital images and gradually merges them to complete pages. The automation degree of this process is determined by the condition of the fragments and the type and degree of destruction: While paper torn to larger, fewer pieces with no further damage done to the resulting fragments can be reconstructed in a largely automated process, other patterns of damage such as very many and very small pieces or additional damage at the edges of the pieces call for more human interaction in the reconstruction process.

Post-production applied to the reconstructed documents can also vary. In some cases, it may suffice to virtually reconstruct single pages. Often it may also be necessary to perform image editing on the reconstructed documents to enhance legibility. In specific cases, automated content interpretation and analysis are needed to tackle larger amounts of source material. Thus, this is an essential part in current research.



## DIGITIZATION



## NEXT LEVEL SCANNING

Digitizing paper fragments for virtual reconstruction turned out to be challenging. In the early stages of ePuzzler development researchers assumed they could work with a commercially available document scanner. But it quickly became evident that no scanner on the market could meet all the requirements of automated virtual reconstruction.

The demands on the digitization process are enormous: Fragments of any size must be scanned from both sides without reflections and shadows, true to the original's scale and color and with stable resolution across the entire digital image. Otherwise, relationships between pieces, such as identical pigmentation or neighbored edges, cannot be identified. Therefore, color deviations of the scanner's lighting source and shape inaccuracies produced by mechanical movement irregularities must be eliminated. Fragments must be digitized on a background they can easily be separated from, similarly to blue-screen-techniques. In order to deal with any given number of fragments in a justifiable timeframe, feeding and image capturing have to happen at a high throughput rate. Once reconstruction activities moved to more fragmented and fragile documents, the team was further challenged with passing fragments through a digitization unit without inflicting further damage to the pieces.

In the end, it became evident that a new and tailored approach for digitization was required. In cooperation with MFB MusterFabrik Berlin GmbH, Fraunhofer IPK has developed a new type of document scanner that meets all these demands. It supports high-precision, fast scanning of nearly any fragment size, ensuring digital images are an accurate reflection of the original's shape and color. Additionally, the device captures transmitted light images that eliminate the need for bluescreen masking. During scanning, the fragments are fixed by a specimen holder consisting of two glass plates which shields them against pressure and friction. Cultural institutions worldwide have expressed high interest in this scanner, not only for reconstruction purposes, but also to digitize fragile documents to guard their content against decay.



Very often, damaged documents cannot be put on a scanner straight away. Their pages have to be separated, straightened or wedged fragments have to be carefully parted from the rest of a document before digitization can take place.



### ePUZZLER AND MORE

The core of virtual reconstruction is the »ePuzzler«, a software which uses complex image processing and pattern recognition algorithms to construct complete pages from digitized fragments. In addition, it provides tools to examine and correct questionable or ambiguous puzzle results.

ePuzzler's methodology is comparable to jigsaw puzzling: The software initially calculates different features of the fragments such as outlines, paper color, writing, or ruling. Similar fragments are automatically clustered into subsets by means of intelligent search



space reduction. The actual reconstruction takes place within these reduced quantities. Corresponding features are investigated along the snippet contours. If two fragments match, they are digitally glued and viewed as a larger fragment during further reconstruction. In 2013, the technology was awarded the EARTO Innovation Prize.

But in many cases, it doesn't even suffice to reconstruct many single pages. Especially when writing materials have been subject to decay, contamination or water exposure, it can be necessary to digitally enhance the images to promote legibility. In these cases, a future solution may be to apply additional light spectrums during the digitization process. After reconstruction, automated pattern recognition processes separate foreground and background information or improve the contrast between text and background. This can be used to overlay the damaged original with reconstructed writing.

To support automated forming of multi-page documents and archive files from reconstructed pages, an extension of the virtual reconstruction technology will eventually be content analysis. Researchers are currently engaged in identifying writers by their unique handwriting, as well as detecting and reading watermarks, stamps and more. To this end, the corresponding image patterns are structurally analyzed with machine learning algorithms such as deep learning and neuronal networks.

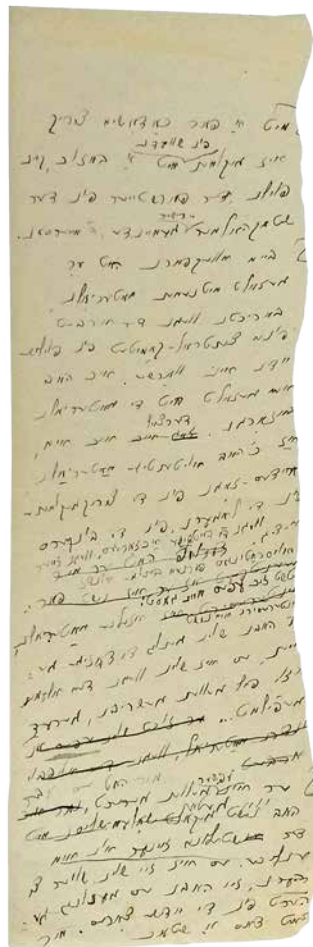
↘A major difference between virtual reconstruction and a jigsaw puzzle: In many fragmented documents, pieces are missing. Reconstruction has to find ways to deal with these gaps.

## NOT JUST RIPPED PAPER

A current Fraunhofer IPK reconstruction project is restoring files from the Asociación Mutual Israelita Argentina (AMIA) building in Buenos Aires, the center of the Jewish community in Argentina. The building was destroyed in a bombing on 18 July 1994, killing 85 people and injuring 300. Over 400 nearby homes and businesses were destroyed or damaged. The bombing also affected the archives and library of the Fundación IWO, Idisher Visnshaftlejer Institut – Instituto Judío de Investigaciones. This research institution aims to enhance knowledge of the history of Jewish life, with a particular focus on the history and culture of Eastern European Jews. The organization's archives and library are among the world's most important resources for research into Jewish life and Jewish history in Europe, the Yiddish language, literature and folklore, the various aspects of the Shoa and the history of Jewish migration to America.

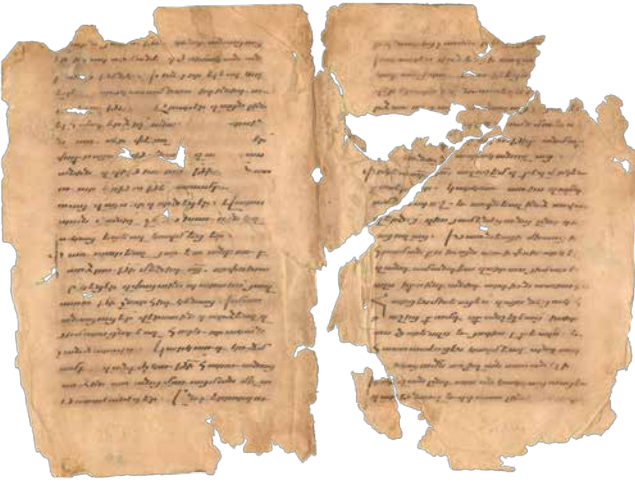
IWO guards a unique collection of documents and books in Yiddish, as well as historical documents relating to Jewish history in Argentina and many publications, documents and other Jewish cultural artefacts rescued from the National Socialists and removed from Europe. These documents are a prime example that virtual reconstruction's significance goes far beyond piecing shredded paper back together. The IWO archives contain personal documents of Jewish personalities who may not have been »big names«, but nonetheless were important figures within the Jewish community or left notes that may develop great importance for tracing a certain historic topic.

One such lesser-known character is Tania Fuks, whose handwritten legacy is kept at IWO. A refugee herself, after World War II she toured the displaced persons' camps and wrote a detailed report on their situation. Some of these notes were published in the early fifties, the rest remain unpublished. Tania's manuscript notebooks were severely damaged in the 1994 blast – an eloquent witness to the loss of information humanity suffers when archive files are destroyed.



TANYE (TANIA) FUKS (1896 – 1950)

After starting her writing career in 1922, Jewish journalist Tania Fuks wrote articles for different Jewish newspapers, mostly in Yiddish. She suffered deportation, survived a ghetto and a camp, and lived in different displaced persons' camps in Germany after the war. Via France and Uruguay, eventually she came to Argentina. Apart from journalistic features, her works include translations and a book on her wartime experience. From her position as cultural director of the Yiddish division of AMIA, she was a central figure in the cultural reconstruction of Jewish life after the war.



## FIGHTING DECAY IN ARCHIVES AROUND THE WORLD

Supporting reconstruction of IWO documents is an excellent example of Fraunhofer IPK's work in the restoration of cultural assets, but it is just one of many activities. A current project with Matenadaran – Mesrop Mashtots Research Institute of Ancient Manuscripts in Armenia shows that reasons other than physical force such as human violence or natural catastrophes can lead an organization to seek reconstruction expertise at Fraunhofer IPK.

With around 23,000 manuscripts and 30,000 other documents written in Armenian, Persian, Arabian and other languages, Matenadaran guards one of the world's largest collections of medieval manuscripts and books. Even though Matenadaran's manuscript collection is under permanent archival and conservational care, recent studies have shown that parts of the collection are in a remarkably poor physical condition. Restoration and physical reconstruction of damaged artefacts are key tasks confronting Matenadaran. In this context, a specially challenging object is one of the first surviving copies of St. Gregory of Narek's Book of Prayers, produced between the 13th and 14th century. To this day, the

Narek prayer book is used by Armenian Christians. Since the original version of this book from the 10th century is lost, great importance is attributed to conserving the early copy.

Restorers are facing the problem that the manuscript is already highly fragmented, plus individual pieces have been disarrayed. Therefore, Fraunhofer IPK was mandated with generating digital fragment images fit for reconstruction as well as experimentally performing virtual reconstruction of the prayer book pages. The virtual reconstruction will then be used as a template for physical reconstruction. Since the fragments are also highly damaged along their edges, this seemed like an impossible task. Yet Fraunhofer IPK experts could successfully prove feasibility, opening up new possibilities for fighting written material decay in archives and libraries around the globe.



Document reconstruction is an expensive and time-consuming task. But it is worth the effort, if it can help conserve valuable cultural assets for future generations.

## CONTACT

Dr.-Ing. Bertram Nickolay  
Phone: +49 30 39006-201  
bertram.nickolay@ipk.fraunhofer.de

Dipl.-Ing. Jan Schneider  
Phone: +49 30 39006-203  
jan.schneider@ipk.fraunhofer.de

## IMPRINT

Fraunhofer Institute for Production Systems  
and Design Technology IPK  
Pascalstr. 8-9  
10587 Berlin

Phone: +49 30 39006-0  
Fax: +49 30 39110-17  
info@ipk.fraunhofer.de  
www.ipk.fraunhofer.de

## Publisher

Prof. Dr. h. c. Dr.-Ing. Eckart Uhlmann

## Editor

Katharina Strohmeier  
Jan Schneider

## Layout

Andy King

## Images

Asociación Mutual Israelita Argentina (AIMA) (1, 2, 4r, 5, 6l, 7r)  
Fraunhofer IPK/Katharina Strohmeier (4l)  
»Marc Turkow« Center at AMIA (6r)  
Matenadaran – Mesrop Mashtots Research Institute of Ancient Manuscripts (7l)

## PARTNERS

MUSTER  
FABRIK  
BERLIN .

The projects with Fundación IWO and  
Matenadaran described in this brochure are  
funded by the German Federal Foreign Office.



Auswärtiges Amt