A new Device for Automated Scanning of Books

Arved Carl Hübler, Günter Seidl*

Introduction: Digitisation of Books

More than 10 million titles of books were published in the past. Every year 120,000 titles in English, 90,000 in German and many titles in a lot of other languages are added, but only a small share of this in a generic digital format.
The production chain may be digital, the publisher, however, does not take care of the access to this information in later times.
Today: There is a significant demand in reprinting old titles, but the publishers kept only sometimes the films and nearly never usable data.
Future: There will be a huge demand in making old titles digitally available.
Nowadays, the dream of having a total access and a digital backup for all knowledge and cultural heritage of mankind is a question of technology and costs.
Digitisation of books becomes an important subject.

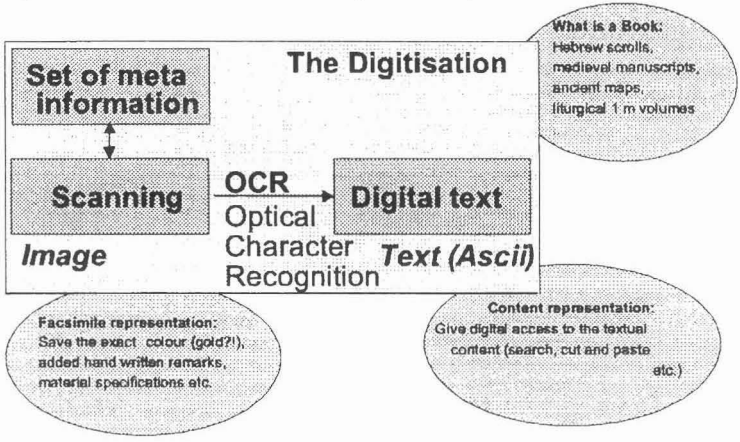

Fig.1: General Problems in Digitisation of Books

* Institute for Print- und Media Technology, Chemnitz Technical University; http://www.tu-chemnitz.de/pm

1. State of Technology

The digitisation of older analogous library catalogues is done for several reasons. An outstanding example for this is the digitisation of the Gutenberg-Bible at the Digitisation Center of the State and University library in Göttingen. It helps to research objects without using or damaging the original. Cultural possessions are preserved and can be used nevertheless. In this way, books that cannot be lent are open to wide markets.
„Subito", a document supplier of essays and books for German libraries, is an example for a service that makes possible to recherche and order on-line and that delivers the needed documents via E-mail directly to the user's workplace. The prerequisite for this, however, is that the literature has to be taken from the shelf and digitised by hand at the bookscanner.
In Retrospective Digitisation depending on the copy it is differentiated whether the data record is to be more image or text oriented, whether an as faithful as possible copy of a facsimile with high resolution and exact coloration is to be produced, or whether it is supposed to be a data record that has to allow content recognition as well as the further handling of the text (OCR).
The objects being digitised from the library catalogue could be:

- books,
- loose sheets,
- cards,
- photographs,
- microfiches.

The Task for this Project: There are a lot of projects in digitising old books. This project looks for the "bread & butter" digitising jobs:

- content oriented representation requested
- not the most precious books
- standard book formats

The goal is an automated digitising process for books to achieve low digitising costs per title. The first step is developing and building an automated scanning device.

Three main scanning technologies are available today:

1. Flat-bet scanner: Scanning the page surfaces of an opened book. To scanall pages, a complicated manual handling is necessary to turn the page. There is no automated system known. No real solution is thinkable. In order to avoid this manual work, at the moment a different method for the digitisation of books is used. A method, in which the books are unsewed and loose sheets

are fed to the scanner is not acceptable for valuable books that should not be destroyed.

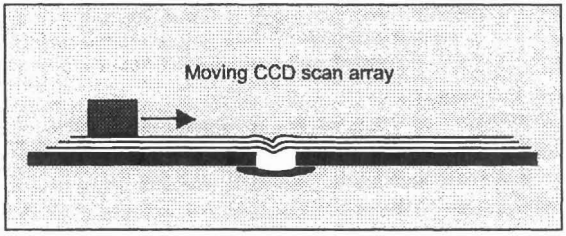

Fig.2: Many standard flat-bed systems available

2. Drum scanner: An application is not possible.
3. 3D-scanner (Digital camera): Scanning the page surfaces of an opened book. To scan all pages, a manual turn of the page is possible. There is no automated system known. An automated system seems to be possible. For the scanning of books 3D scanners, in general called bookscanners, are used. In the course of this the book, with the open pages up, is placed on an even surface. The camera is hung above it.

Cameras that move one CCD line of about 5,000 pixels across the area to be scanned reach in the camera a resolution of 300 dpi in A2 format. This amounts to ca. 17.5 Mio pixels/image. (Zeutschel, 1999). An example for an available system is the Omniscan 6000 Color by Zeutschel, Tübingen/Hirschau. A higher resolution in smaller formats is reached by a camera that is height-adjustable and that can be brought into line with the format. High-End devices, such as those used for the digitisation of facsimiles, have a CCD-area-sensor with 8000 x 9700 pixels, which comes to ca. 77.6 Mio pixels/image (Liebetruth, 1999).

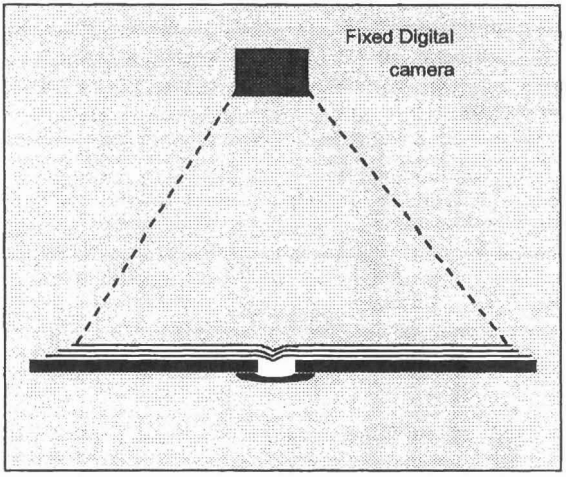

Fig.3: 3D-Scanningsystem: (Special suppliers: Minolta, Xerox, Zeutschel a.o.)

Furthermore, according to purpose, it is differentiated whether the camera is designed for black and white, gray values, or color shots.
All bookscanners on the market require the turn-over of the pages by hand. The technological development of these bookscanners includes higher resolutions and faster data processing, but still an automation of turning over the pages is not known. Therefore, the performance will always be limited by the handling of the books. The crucial disadvantage of these bookscanners is that human labor is necessary during the entire digitalization process. This task is very monotonous and demands extreme concentration to receive a faultless file.
The Important Automation Feature: Turning the Page
If one does not want to destroy the book, for automation a mechanical system for turning the pages is needed. Advanced paper handling is not the standard knowledge of digital scanner suppliers, therefore our project was created to use experiences from paper handling in printing presses to solve the problem.

2. Demands for Scanning Devices

The system parameters for capturing data are part of the system of the camera and data processing in the computer. This difficulty is not issue of this research project.

For the concept of the system the size of the books which may be digitised is very important. In order to establish the bounds regarding book formats, inquiries were held in the reading room of the library of the Chemnitz Technical University.

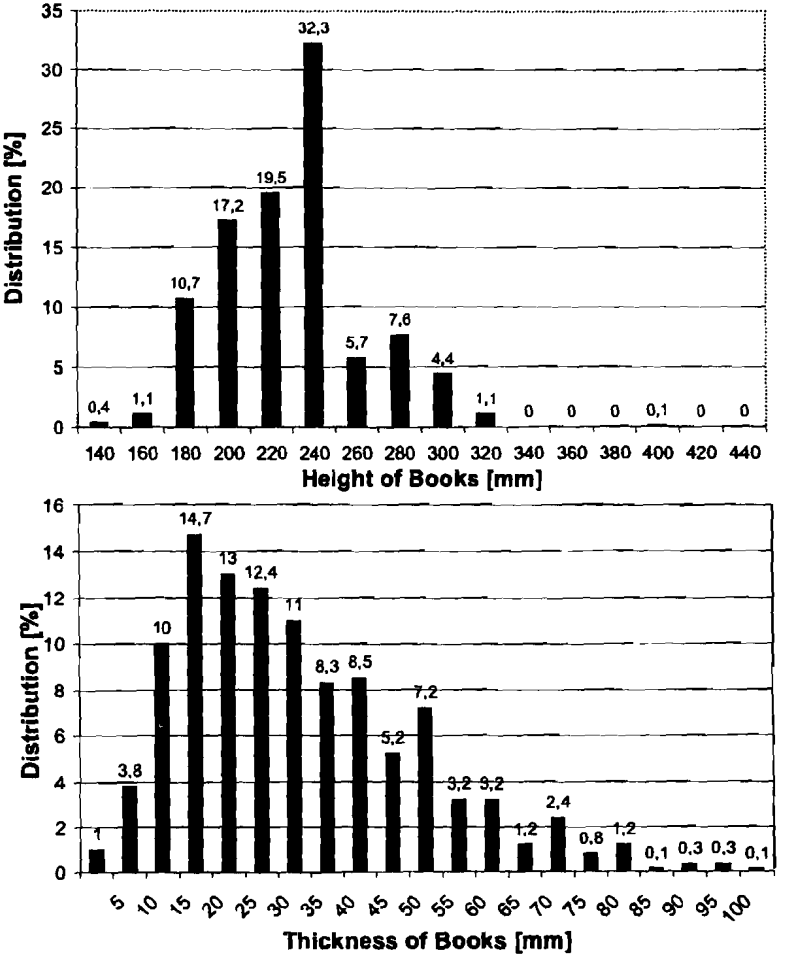

Fig.4: Analysis of 3,009 Books in the Library of the University
[Subjects: Biology, Mathematics, Technics, Physics, Chemistry/Pharmacy, Economics, Law, Education Science, History of Art, English Philology, History, Bibliographic Reference Books, German Philology, Philosophy, Musicology, Politics]

Two thirds of the measured books have a height ranging from 200 – 240 mm. An insignificant number of books has a height of over 300 mm. For the prototype, this allows to determine a maximum format for an open book of 300 x 450 mm (corresponds with A2). With this about 98 % of all books of the selected subjects are included. Large-format geographical books were deliberately excluded. It will be proven if a larger special model is needed in the future.
The analysis of the thickness of the books revealed that none of the measured books was thicker than 100 mm. Therefore this measurement is sufficient for the mutual height compensation of the supporting surface for books.
For the time being, the productivity is mainly determined by the scanning time which is ca. 10s/image for A2 format (this corresponds with a double page A3). It takes for instance 25 minutes to digitise a book consisting of 300 pages. Within this time the turning of the page is done manually.

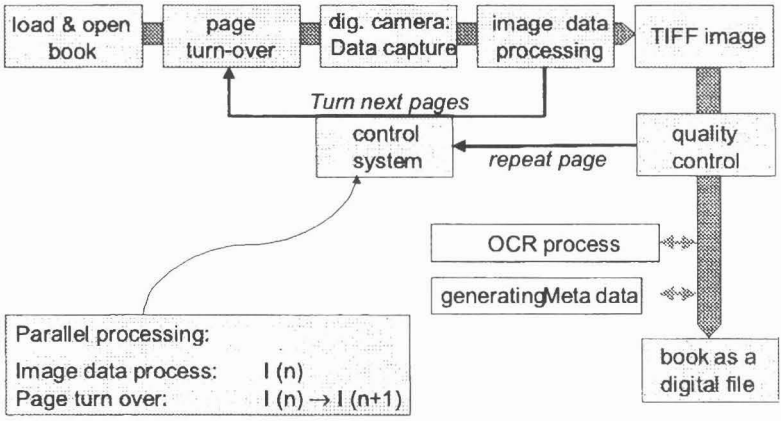

Fig.5: Functions in a Scanning Process

Page Turn-over Devices known for other Applications

In patent literature many patents for turn-over devices can be found. Some of the latest patents in this field are to be specified. One group is qualified as support for disabled persons and for music stands. In this connection the books mostly stand nearly vertically, leaning slightly backwards and the setting in

motion is done by foot. Various elements to catch and turn over the pages are proposed: rotating drums with pockets in which the loose sheets are inserted (Moritz, Adolf; 28.01.99), (Moritz, Adolf; 20.02.98), a drum with horizontal axis wherein after use the loose sheets fall into a pocket underneath (Strauß, Gerhard; 18.01.96), a swivel arm with a clamping gripper that rotates around a nearly vertical axis turns the page (Heise, Rudolf; 05.12.94), (Trauschke, Siegfried; 16.03.94).

To imprint into passports or similar small-format thin booklets the pages are shifted by a distributor drum and the forming loop is turned over (Maekawa, Hideki; 04.12.96), (Trauschke, Siegfried; 16.03.94).

For thicker books the page is lifted a bit and runs into a slit. When passing the middle of the book the page is turned over and runs out of the slit again (Maekawa, Hideki; 04.12.96), (Fujioka, Tetsuya u. a.; 01.03.94).

Vertically moving suction cups lift the page to be turned over. The book is moved towards the suction cups and the page turns itself over. (Masaaki, Horikawa; 23.04.92),

For the purpose of turning over in order to digitise, none of the presented solutions is usable. The reasons are the following:

- The first-mentioned patents require a position of the book that is not horizontal.
- Distributor drums could damage the surface of the pages.
- The insertion of the page into a slit involves friction on the page and also possible damages of the surface.
- The book should not be moved while turning the page, but should remain lined-up in position so that the quality of the shooting is not spoiled.

A page turn-over device that is suitable for bookscanners, i.e. digitisation devices with the camera hung above, have not been found in the patent literature yet.
For our requirements, none of these solutions is usable.

Finally Chosen Features for the Design

In general:

- Not a small desktop solution but a solid machinery (e.g. using vacuum)
- The necessary skill for operation has to be as low as possible.

- Focussing on a working solution, but having later system costs in mind

In detail:

- No risk of damaging the surface of the pages:
  insertion of the page into a slit involves friction, no distributor drums, but using suction cups to separate the pages.
- The book should not be moved while turning the page, but remain lined-up in position so that the quality of the shooting is not spoiled.
- With one shot, two opposite pages are digitised.
- A fixed geometrical situation with a fixed resolution, no optical adjustment for different book sizes to optimize the resolution.
- Using covering glass tops to get an even optical plane guaranteeing the quality of the shooting.
- The accessibility from the front and sideways is guaranteed.

3. Realization

After a systematic contemplation of the demands on an automatically operating bookreader with the function to turn pages the following solutions were chosen and specified. The page turn-over device will be a prototype intended for the examination of optimum conditions for turning the pages of books of several formats and paper qualities. Therefore, a wide adjustment area for process relevant parameters is planned.

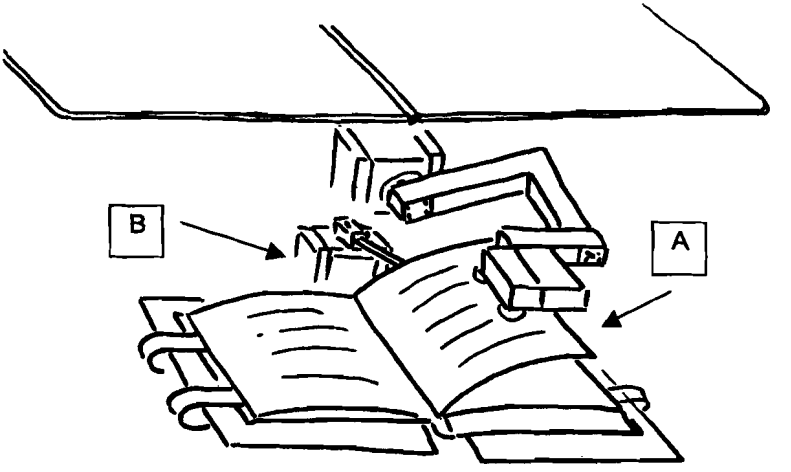

Fig.6: Using Suction Cups and a Turn Over Stick

Function sequence:

1. The book is attached to the base by clips.
2. The lever with the suction cups swings over the page to be turned and sinks down onto the page, it is sucked (A)
3. Suction air on, the page is sucked towards the suction cup.
4. The suction cups tip over and lift the page to be turned (see fig.3).
5. The turn-over device (B) rotates around a vertical axis below the lifted page, suction cups release it.
6. The turn-over device turns the page over.

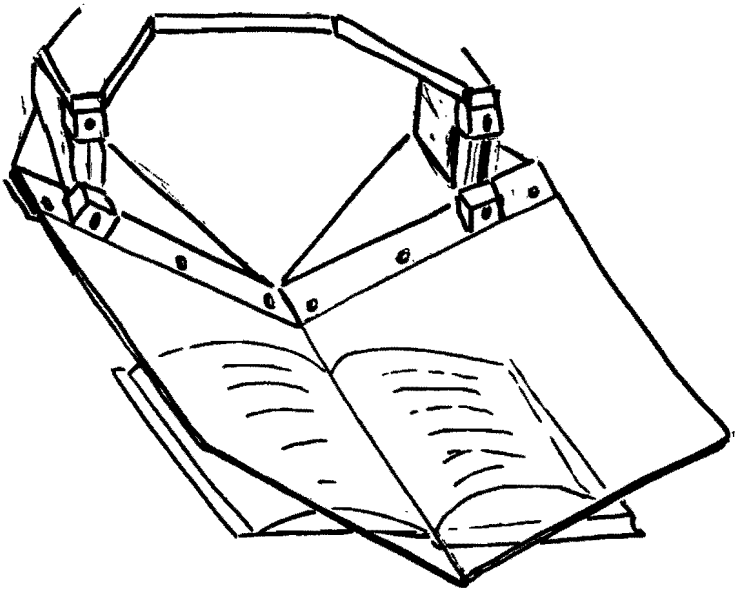

Fig.7: Putting a Divided Glass Top down on the Book

7. The glass top, divided in the middle, sinks down onto the book.
8. The two inner edges of the divided glass top lie on the gutter of the book.
9. During the further sinking the two parts of the glass top flatten out the pages.

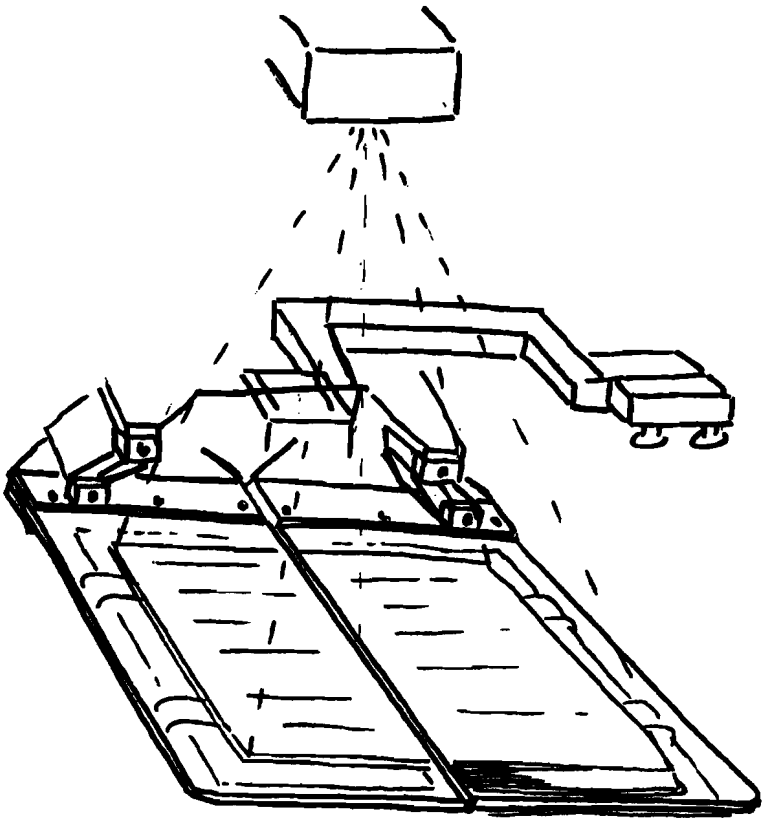

Fig.8: A clear Area for Shooting

10. After the page is turned, the lever with the suction cups rotates around a vertical axis to the right, out of the shooting area.
11. The timing for putting down the divided glass top prevents that the turned page leafs back.
12. While the pages are being turned, the supporting surface for the books adjusts itself to the differences in height.

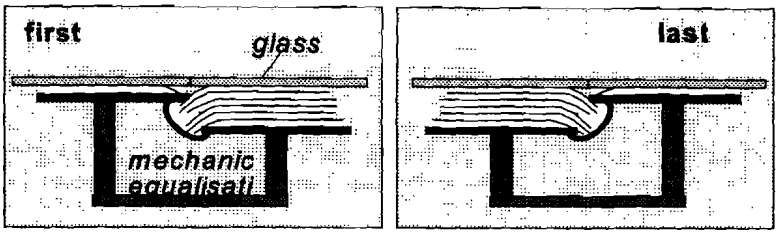

Fig.9: Differences in Height During the Page Turning

Reasons for the Detail Solutions:

The book's divided supporting surface (book seesaw) is designed that way that the opened book lies with the opening on top. The camera is hung above. This arrangement allows:

- A free movement of the suction cups to catch, release, and lift the pages of the book.
- According to the changing number of pages throughout the process of turning the pages, the height of the divided supporting surface adjusts to the book so that there is always an equal level of the open pages.
- A visual control of the trouble-free operation is possible because of the pages being visible.

The below atmospheric pressure within the suction cup is generated by the blow air and by component parts that work on the ejector principle. Advantages:

- An additional vacuum pump is not necessary.
- The component is very small. This way it can be placed directly at the suction cup.
- The below atmospheric pressure and with it the suction output can be controlled by reduction.

The movement of the suction cups is done by pneumatic rotation and lift cylinders. Advantages:

- The parts are prefabricated, with it quality is guaranteed.

- Simple adjustments concerning swing angle and lift length are possible. This results in constant and reproducible conditions while working at a book and less expenditure in adjusting formats.

The lifting and turning over of a page will be realized by divided component parts. Advantages:

- Shorter distance of the suction cups.
- A turn-over device with swing movements around a vertical axis above the middle of the book takes the lifted page and puts it on the other side.
- After having turned over the turn-over device lies on top of the page and prevents that stiff pages leaf back.
- The suction cup at the swing lever can swing out of the book area during the shooting.

After the page is turned over, a divided glass top sinks down onto the book in order to receive a flat shooting area. Advantages:

- The glass top that is divided in the middle sinks down on the gutter and flattens out the pages towards the edges.
- The covering glass tops result in an even optical plane guaranteeing the quality of the optical shooting.

The component elements to move the suction cups, the turn-over device, and the divided glass tops are arranged behind the open book. Advantages:

- The accessibility from the front and sideways is guaranteed. This way the bookreader is user-friendly.
- It facilitates the subsequent extension of the bookreader with a magazine feeding containing several prepared books.

4. The Problem of Separating Single Pages

For turning the pages reliably, because only this way a faultless file can be guaranteed, the issue of the influence coefficients from paper properties is interesting.
A wide range of paper properties could appear:

- Large variety of paper grades in the different books
- The paper may be old, dirty and torn

Using suction cups, there is the danger that not only one but two or more pages are caught

- The paper edges stick together. In order to avoid this, the book is loosened before it is inserted into the system.
- The air permeability of paper is so high that the following or even more pages are drawn in.

One possibility to part the second drawn in page is to tilt the suction cup.
Due to the paper's elastic bending strength **E** the page comes off if the bending force is stronger than the suction force as function of air permeability

Bending strength cross-grain, expressed by the E-module,
calculated according to

$$E = \frac{Q \cdot l}{8 \cdot I \cdot f}$$

in which are
Q in [g] the net weight of the paper strip, clamped at one end only,
l in [cm] the free clamping length of the paper strip,
I in [cm^4] the 2nd degree area factor and
f in [cm] the bending.

Fig.10: Wide Range of Paper's Elastic Bending Strength **E**

The tested paper grades are:

Grade	Grammage [g/m²]	Thickness [mm]	Air Permeability [ml/min]	E-modul [N/mm²]
1 Telephone book paper, woodpulp, slightly glazed	37	0,054	400	121 x 10^3
2 Gravure paper from a booklet, woodpulp, glazed	61	0,066	60	322 x 10^3
3 Offset paper from a technical book, woodpulp, uncoated paper	61	0,068	58	108 x 10^3
4 Offset paper from a booklet, wood-free, uncoated paper	80	0,103	833	238 x 10^3
5 Offset paper from a technical book, slight woodpulp, slightly glazed	104	0,133	17	392 x 10^3
6 Art paper from a coffee-table book, coated on both sides, wood-free, matt	135	0,122	Measurement impossible *)	392 x 10^3
7 Cardboard, blank, white, wood-free, uncoated surface	187	0,215	500	257 x 10^3
8 Cardboard, Couch Certificate, chamois, woodpulp, glazed, uncoated surface	193	0,213	40	339 x 10^3

Table 1: Summary of the Paper Parameters
*) 100 ml/10min at 250mm underpressure watercolumn

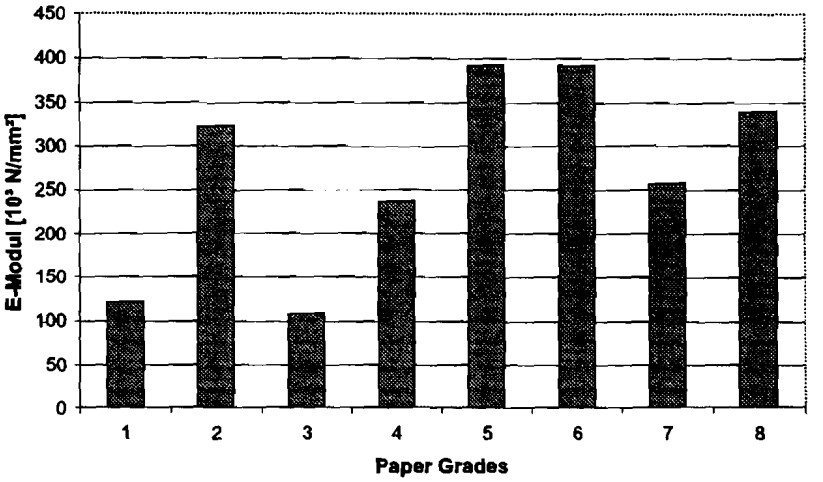

Fig.11: Wide Range of Paper's Elastic Bending Strength E

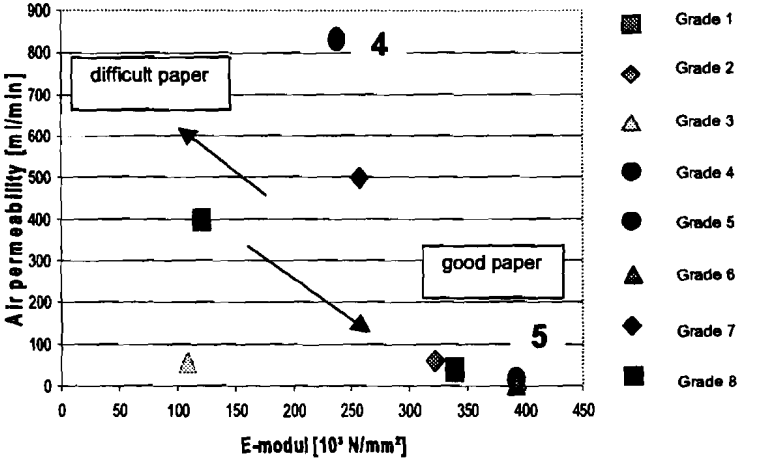

Fig.12: Air Permeability vs. Elastic Bending Strength: Good and Difficult Paper Grades

Comments on the Measurements

The combination of low air permeability and high bending strength is advantageous for turning the pages reliably. Art papers that are coated on both sides belong to this group. By all appearances, these are easily classified as such.
More critical, for example, is the difference between the offset papers grade 4 and grade 5. These are both uncoated papers, wood-free and nearly wood-free. There are hardly any differences in the optical (reflected light) and haptic properties of both. The air permeability of grade 5 is 17 ml and that of grade 4 830 ml. A high asset of air permeability seems to be disadvantageous for a reliable turn-over.
However, both paper grades show one remarkable difference: clouds in through-light. This means an uneven paper formation, which is more uneven within the air permeable paper than within the dense paper. Further tests are necessary to check if there is a correlation between the clouds and air permeability. This would be a simple solution to judge whether the paper is a problem for the page turn-over or not.
The adjustment possibilities to receive optimum conditions for the page turn-over at the bookreader are:

- the negative air pressure that is adjusted to the paper grade
- the swing angle and the swing radius when the suction cups move and
- an appropriate suction cup (shape, size, smoothness)

The goal is, on the one hand is, an optimum adjustment that allows to turn pages of a wide range of different paper grades reliably and, on the other hand, sound hints how paper problems need to be solved.

5. Conclusions

- We have found a solution for an automated book scanning device.
- The goal for this project is to realize a prototype system: It will run until the end of 2000.
- Some practical experiences are necessary to optimize the paper related adjustments for separating single pages.
- A commercial realization by a supplier of book scanning technology may follow.

Advanced work:
- Developing a quality control an a error detection technology.
- Developing a system for an automated change of book volumes.

Utilized literature:

Zeutschel
1999 Brochure of the Zeutschel GmbH, D-72070 Tübingen

Liebetruth
1999 High-resolution digital camera "Picture Gate 8000"
http://www.sub.uni-goettingen.de

Lippsmeier
1999 "Friedrich-Tabellenbuch Metall- und Maschinentechnik" Dümmler Verlag Bonn, pp 2-20

Investigated patents:

Moritz, Adolf
28.01.99 „Vorrichtung zum Wenden von Blättern"
EP 0937 583 A2

Moritz, Adolf
20.02.98 „Vorrichtung zum Wenden von Notenblättern"
DE 298 02 973 U1

Maekawa, Hideki
04.12.96 „Umblättervorrichtung"
DE 196 50 312 A1

Strauß, Gerhard
18.01.96 „Maschine zum Umblättern loser Seiten"
296 00 791 U1

Heise, Rudolf
05.12.94 „Vorrichtung zum Umblättern der Seiten eines Buches oder dergleichen"
DE 44 43 221 A1

Trauschke, Siegfried
16.03.94 „Blattwender"
DE 94 04 375,

Fujioka, Tetsuya u. a.
01.03. 94 „Bilderzeugeeinrichtung mit Buchlese- und Seitenumblätter-Funktionen"
DE 44 06 667 A1

Kabushiki, Kaisha, Toshiba
28.07.92 „Apparatus for printing images on booklets"
5 183 347

Masaaki, Horikawa
23.04.92 „Apparatus for turning over Page",
5 - 30 14 87 (A),

Thanks
We thank DFG (German Research Society) for supporting the work financially. *[DFG Project No. III N2-54281 (1)]*