Image recognition and image similarity

A different approach for accessing large scale digital collections

Project background

- Numerous digitization projects since 1997
  - Manuscripts, incunabula, newspapers, books
  - Up to 1 M objects from 8th – 20th century

- Deep indexing available only for a part of the books
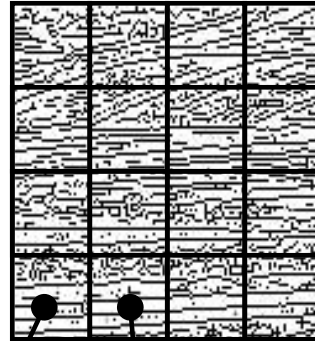
→ Where are the images?

Seeking for a different approach: The image-based similarity search

- 2011: Start of cooperation with the *Fraunhofer Heinrich Hertz Institute Berlin (core technology)*
  - Based upon existing expertise of Fraunhofer HHI
  - Initiated by Dr. Markus Brantl (head of Dig. Library Dept.)
- 2012: First prototypical implementation (250 volumes)
  - Optimization for BSB's holdings
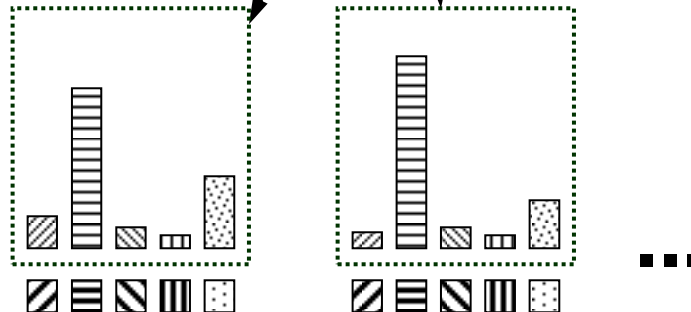 - Development of API and web interface
- Online since April 2013

How it works

How it works: Edge detection and description

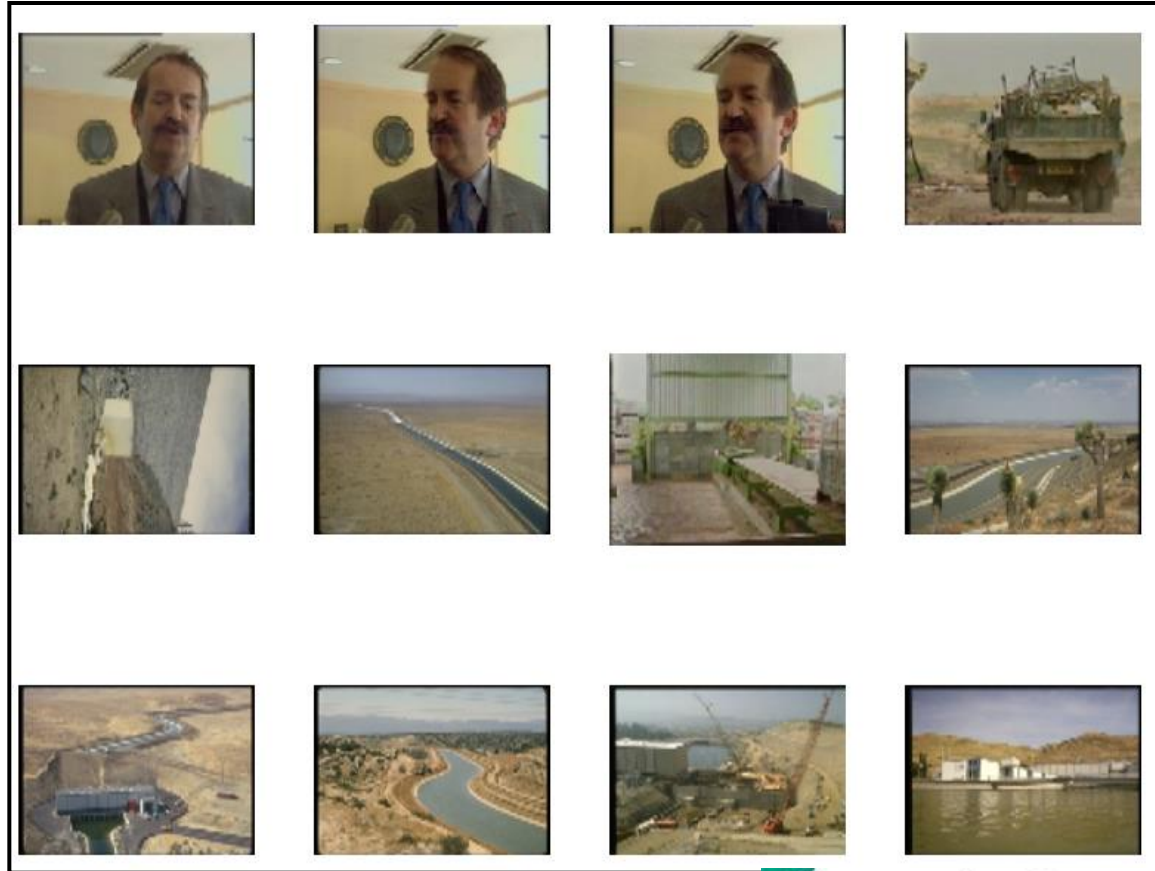Edges, ie transitions between areas of different brightness

$\rightarrow$ egde histogram
  $\rightarrow$ image descriptor

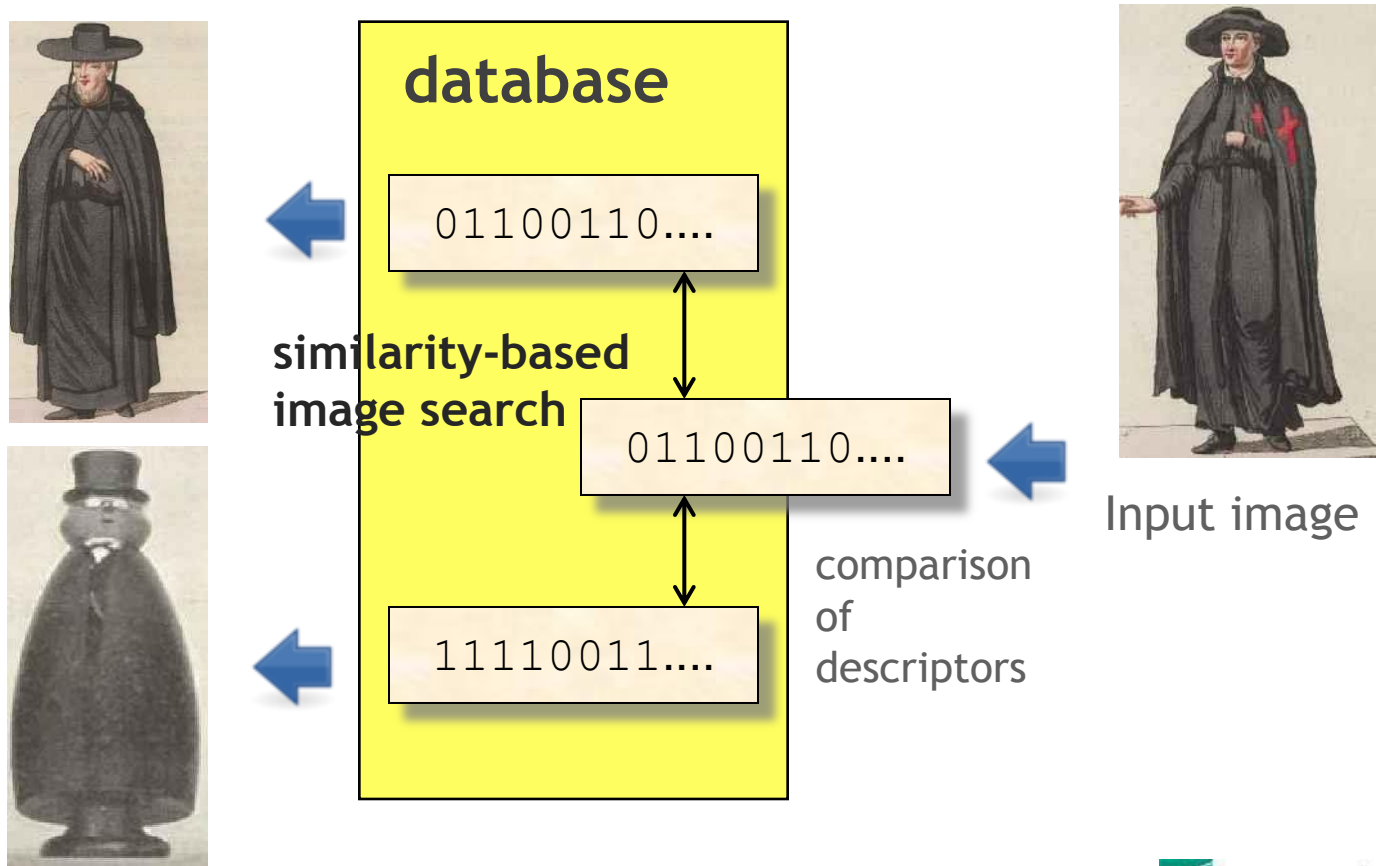# How it works: Color distribution

Color distribution

How it works

1. *One-time* processing of each image
   - Analyzing properties
   - Generation of a unique descriptor
   - Filling an index of descriptors

2. Index is loaded into RAM permanently
   - High availability for the application
   - Fast search among millions of images

BSB Bayerische StaatsBibliothek
Information in erster Linie

MDZ

The similarity search in practice

database

01100110....

similarity-based image search

01100110....

11110011....

comparison of descriptors

Input image

Search results

Additional challenges with BSB's holdings

- Recognition of images inside of pages
- Image/text separation

Additional challenges with BSB's holdings

- Recognition of multiple images per page (segmentation)

Additional challenges with BSB's holdings

- Recognition of book illumination, initials, drawigs, woodcuts, copperplate engravings, old newspaper photographs, ...

Recognition accuracy
Some examples

Recognition accuracy, example #1

Similar body shape, dress colour and foot position

Recognition accuracy, example #2

Coat of arms: similiar shape, similiar background – and a striking (red) tongue

Bayerische
StaatsBibliothek
Information in erster Linie

BSB

MDZ

Recognition accuracy, example #3

The same printing plate used for different books, with different coloration

Recognition accuracy, example #3

Recognition accuracy, example #4

Recognition accuracy, example #5

Church interiour and train station halls - similar shapes & lighting conditions
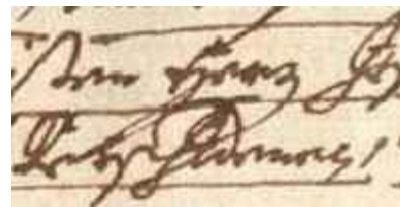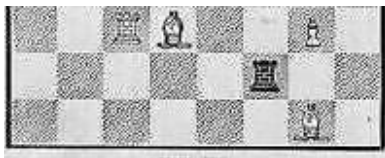
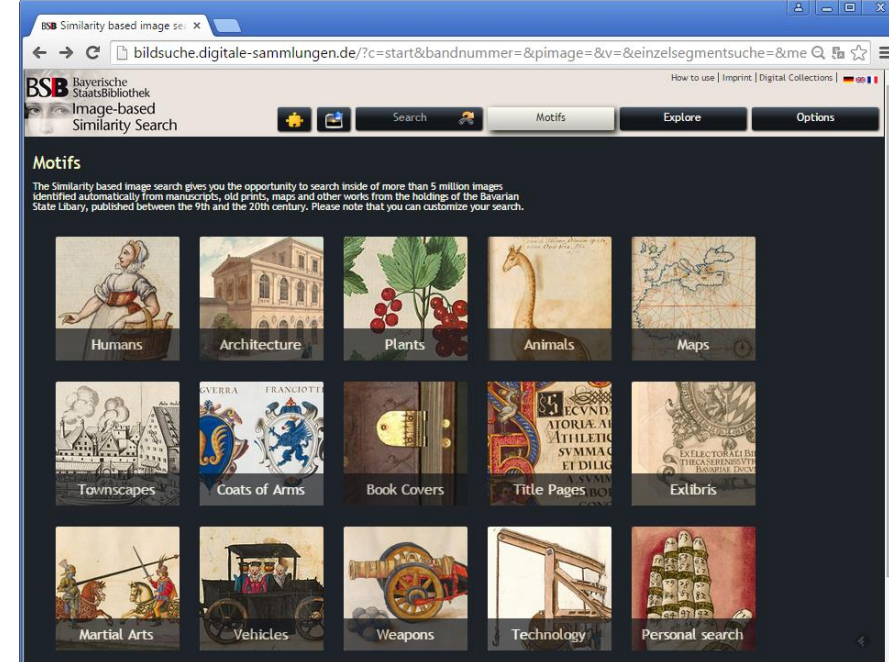# Recognition accuracy, example #6

Four legs on the ground …

The limits of image recognition

- Stains & ripped pages
- Decorative headlines
- Exceptional shapes
- Table Works
- Newspaper ads
- Sheet music
- …

Web application – Functionalities overview

- Explorative access
- Customizable Simple Search
- Skimming of books
- All images from one book
- Compare your own image – Upload

BSB Bayerische StaatsBibliothek

Image-based Similarity Search

How to use | Imprint | Digital Collections |

Search   Motifs   Explore   Options

Motifs

The Similarity based image search gives you the opportunity to search inside of more than 5 million images identified automatically from manuscripts, old prints, maps and other works from the holdings of the Bavarian State Libary, published between the 9th and the 20th century. Please note that you can customize your search.

Humans
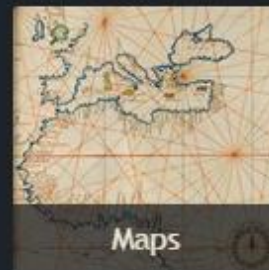
Architecture

Plants

Animals

Maps

Townscapes

Coats of Arms

Book Covers

Title Pages

Exlibris

Martial Arts

Vehicles

Weapons

Technology

Personal search

How to use | Imprint | Digital Collections |

BSB Bayerische StaatsBibliothek

Image-based Similarity Search

Search

Motifs

The Similarity based image search gives you the opportunity to search inside of more than 5 million images identified automatically from manuscripts, old prints, maps and other works from the holdings of the Bavarian State Libary, published between the 9th and the 20th century. Please note that you can customize your search
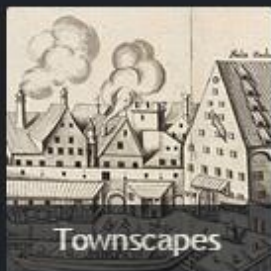
Humans

Architecture

Plants

Townscapes

Coats of Arms

Book Covers

Martial Arts

Vehicles

Weapons

Options

Time frame

1013 - 1945

Color Edge Ratio: This controller weights the color edge ratio of the search. A value of 0.5 considers colors and edges equally.
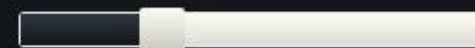
0.55

Edges   Colors

Minimum similarity: This controller sets the minimum similarity. The higher the value, the higher the similarity, while 1.0 would be congruence. It is recommended to set a value of at least 0.85 but not higher than 0.95.
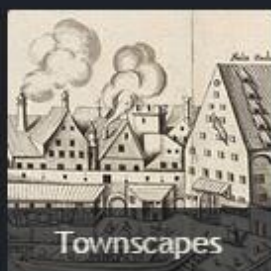
Minimum similarity 0.85

Results: Maximum number of results to be displayed

50   Results

Reset to default

bildsuche.digitale-sammlungen.de/index.html?c=start&l=en

BSB Bayerische
StaatsBibliothek

Image-based
Similarity Search

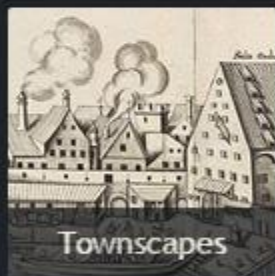Bildnis   Motifs   Explore   Options

Motifs

The Similarity based image search gives you the opportunity to search insid...
identified automatically from manuscripts, old prints, maps and other wor...
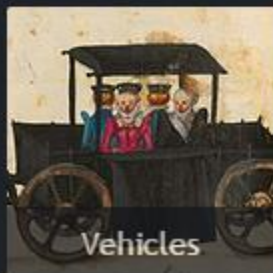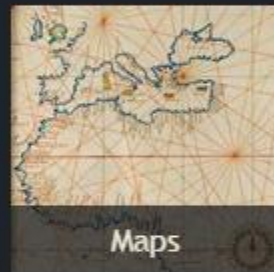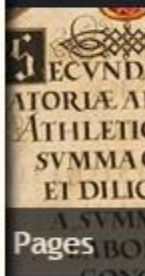State Libary, published between the 9th and the 20th century. Please note...

Nr. 47 BILDNIS des Ritters Jörg
von Freundsberg [...]
[1894]

Tafel I. Einband für Albrecht V.
mit seinem BILDNISse. [...]
[1917]

BILDNIS Johanns V. oder Johanns
VI., Kurfürst von Trier [...]
[1554]

Künstlerportraits. BILDNISse
italienischer Künstler und einiger
Gelehrter [...]
[1624]

19. Das BILDNIS Kurt Eisners
[1919]

BILDNISse der Vorfahren des
Nikolaus Still. Beginnend mit
Wilhelm [...]
[1575]

Nr. 7 BILDNIS des Memminger
Reformators Christoph
Schappeler [...]
[1894]

BILDNISse der ersten zwölf
römischen Kaiser [...]
[1624]

BILDNISse
[1939]

BILDNIS, wie die Fürsten
Osterreich ihr Lehen empfangen
[...]
[1555]

BILDNIS des Hermann von Wied,
Kurfürst von Köln [...]
[1554]

Widmungsseite mit
BILDNISmedaillon König Ludwigs
XIV. von Frankreich [...]
[1690]

Nr. 395 BILDNIS Ulrichs von
Schellenberg
[1883]

BILDNISse

Humans   **Architecture**   ...mals   **Maps**

Townscapes   **Coats of Arms**   **Pages**   **Exlibris**

Martial Arts   **Vehicles**   ...nology   **Personal search**

BSB Similarity based image search ✕

◀ ▶ C 🔒 bildsuche.digitale-sammlungen.de/index.html?c=suche_sim&bandnummer=bsb00001649&pimage=00080&einzelsegmentsuche ☆ ≡

BSB Bayerische
StaatsBibliothek

How to use | Imprint | Digital Collections | 🇩🇪 🇬🇧 🇫🇷

Image-based
Similarity Search

🧩 📕 📂   Search 🔍   Motifs   Explore   Options

Search Template: *(37v)*

of: Tirol, Anton:Wappenbuch - BSB Cod.icon. 310 [Cod.icon. 310], [S.l.] Süddeutschland Ende 15. Jh. - 1540

all images of this book

Open Book

Results: 49 Similar images

Image
page 81
from *Tirol,
Anton:Wappenbuch
- BSB Cod.icon. 310
... (Ende 15. Jh. -
1540)*

Image
page 383
from *Sammelband
mehrerer
Wappenbücher -
BSB Cod.icon. 391
... (um 1530)*

Image
page 120
from *Wappen
besonders von
deutschen
Geschlechtern - BSB
Cod.icon. 309 ...
(1475 - 1560)*

Image
page 81
from *Tirol,
Anton:Wappenbuch
- BSB Cod.icon. 310
... (Ende 15. Jh. -
1540)*

Find similar images

Image
page 383
from *Sammelband
mehrerer
Wappenbücher -
BSB Cod.icon. 391
... (um 1530)*

Find similar images

Image
page 120
from *Wappen
besonders von
deutschen
Geschlechtern - BSB
Cod.icon. 309 ...
(1475 - 1560)*

Find similar images

Image
page 67
from *Sammelband
mehrerer
Wappenbücher -
BSB Cod.icon. 391
... (um 1530)*

Find similar images

Flandrische
Stadtwappen
(63v)
page 132
from *Flandria -
Chorographie
Flanderns in
Portraits der
Landesfürsten, den
... (1562)*

Find similar images

Image
page 15
from *Sammelband
mehrerer
Wappenbücher -
BSB Cod.icon. 391
... (um 1530)*

Find similar images

Image
page 14
from *Sammelband
mehrerer
Wappenbücher -
BSB Cod.icon. 391
... (um 1530)*

Image
page 378
from *Sammelband
mehrerer
Wappenbücher -
BSB Cod.icon. 391
... (um 1530)*

Image
page 47
from
*Hofkleiderbuch
(Abbildung und
Beschreibung der
Hof-Livreen) des
Herzogs Wilhelm
... (16. Jh.)*

BSB Bayerische StaatsBibliothek
Image-based
Similarity Search

Search   Motifs   Explore   Options

Personal search

Upload your own image as a template for an image search

Please note:

- use image files of format JPG or JPEG only
- to obtain satisfying results, use a resolution of 150 - 300 ppi
- limit is 1500 KB per file
- timeout of at least 1 minutes after each upload
- the uploaded file will be deleted after the search

Datei auswählen   Keine ausgewählt

Enter security code

1658

Upload Search Template

vionlink PHP UploadS

Version 1.6.8 - HHI V 2.9 [22.04.2013] - 0.85/0.55/ (1013 - 1945) 89

W3C XHTML 1.0

Main benefits of image-based similarity search

- Reliable recognition of images
- Optimized for the specific holdings of the BSB
    - Paintings, woodcuts, copperplate engravings, photographs
    - Color or b/w
- Separation of images from text
- Customizable
- High performance even with large amounts of data

Outlook

Today

- 85.000 digitized works indexed

    $\rightarrow$ 2,878,261 pages

      $\rightarrow$ 5,423,025 images available

Planned for 2015/2016

- 1 Mio. digitized works (complete digital collections)

    $\rightarrow$ 50 M. images assumed

Try it out

http://bildsuche.digitale-sammlungen.de/

http://bildsuche.bavarikon.de/

BSB Bayerische
StaatsBibliothek
Information in erster Linie

MDZ

BSB Bayerische
StaatsBibliothek

Dr. Thomas Wol
Thomas.Wolf@bsb-muenchen.d