



Image-matching technology applied to Fifteenth-century printed book illustration

Matilde Malaspina¹ · Yujie Zhong²

Published online: 20 December 2017

© The Author(s) 2017. This article is an open access publication

Abstract

This article examines how image-matching and content-based image retrieval technologies can be fruitfully applied to track the reuse and circulation of illustrations found in Fifteenth-century printed books. The possibility of tracking precisely and quickly the multiple occurrences of a single woodcut through one or different editions will help scholars to further explore how printed material was exchanged between or copied by different printers, as well as to analyse the working practice of a single printer, through a detailed reconstruction of his use of the illustrations in the composition of the printing sheets. Through the application of innovative computer science technologies, we learn more about the practical use of images, and therefore their cultural and social function, at the pivotal time of the spread of printing in the Western world.

Keywords Image retrieval technologies · Woodcut illustrations · Printing · Fifteenth-century · Book history · Early printing · Machine learning · Computer vision

This contribution is based on the possibility of successfully applying innovative scientific technologies to philological, art historical and bibliographical studies.

The books that were printed all over Europe from the early 1450s to 31 December 1500 are known as *incunabula*. Some 30,000 editions survive today, in about 450,000 copies scattered across more than 4000 public libraries and private collections all over the world.

A full inventory of the surviving books is available in the Incunabula Short Title Catalogue (ISTC) database, coordinated by the British Library (London).¹ Their typographical features are outlined in the Gesamtkatalog der Wiegendrucke (GW).²

The history of each surviving copy (former owners, decoration, binding, manuscript annotations etc., information collectively referred to as *provenance*) has been traditionally described in library catalogues and is now being brought

together in the Material Evidence in Incunabula (MEI) database.³

The texts contained in each edition are being described in detail in the Text-Inc database.⁴

High quality data from MEI is being used in a visualization suite to present the circulation of books over time and space.

These last three digital tools are the product of the 15cBOOKTRADE, a 5-year European Research Council-funded project led by Dr Cristina Dondi and based at the University of Oxford (Faculty of Medieval and Modern Languages/Lincoln College).

Along with types, textual content, and provenance evidence, the other fundamental component of incunabula are

¹ Incunabula Short Title Catalogue is the international database of Fifteenth-century European printing created by the British Library with contributions from institutions worldwide. Each record includes information on author, short title, the language of the text, printer, place and date of printing, and format. Links are provided to online digital facsimiles and to major online catalogues of incunabula (see ISTC at <http://www.bl.uk/catalogues/istc/index.html>).

² See the GW at <http://www.gesamtkatalogderwiegendrucke.de/>.

³ MEI, conceived in 2009 by Cristina Dondi and developed by Alexander Jahnke of Data Conversion Group, University of Gottingen, hosted and maintained by the Consortium of European Research Libraries (CERL), gathers together material evidence from thousands of surviving Fifteenth-century printed books. See MEI at <http://data.cerl.org/mei/>.

⁴ See Text-Inc at <http://textinc.bodleian.ox.ac.uk/>.

✉ Matilde Malaspina
matilde.malaspina@mod-langs.ox.ac.uk
Yujie Zhong
yujie@robots.ox.ac.uk

¹ Lincoln College, Turl Street, Oxford OX1 3DR, UK

² Harris Manchester College, Oxford OX1 3TD, UK

printed illustrations, which in the early stage of printing consisted mainly of the insertion of carved woodblocks into the printing forme.⁵ Hence the name *woodcuts*.

At a time when the spread of printing facilitated the availability of books to wider sections of society, images in books continued to serve as visual aids in deciphering and clarifying the content of the verbal signs, as well as starting points for meditation, memory and thinking, and as a primary intellectual tool in the reading process. From time to time they were used simply to catch the reader's attention, this resulting in a random choice of images, disconnected from the text.

With the introduction of printing, the iconographic apparatus contained in illustrated books gradually started shifting from the illuminated manuscript products, where unique decorations were carried out to fulfil the requirements of a single patron, to multiple copies, mechanically printed and widely distributed, containing illustrations that had to be appreciated and understood by a more general public.

Woodblocks soon became a fundamental part of printers' business capital. On a par with types, paper, and the press itself, they had an economic value: they could be loaned to other printers, they were exchangeable, and they were marketable. And in fact, from the very beginning, many of the woodblocks which had been commissioned and prepared in order to illustrate early Fifteenth-century printed editions started to be copied or re-used in other editions, within the same iconographic cycle, or as single images, illustrating the same or a different text, by the same or by different printers, sometimes in different countries.

Throughout the centuries, many individual efforts have been carried out by scholars in order to explore and better understand how the role of book decoration, and the creative process it entailed, changed with the introduction of printing, and to clarify the relationship between painting, illumination and the different stages of printed production.⁶

In recent years, many projects have also been fruitfully testing the application of digital technologies to different kinds of images, and to early printed images as well.⁷ Nonetheless, a coordinated systematic approach, able to track the

production, circulation, use and reutilization of Fifteenth-century printed woodcuts is still lacking.

In this context, the 15cBOOKTRADE has been working towards the creation of a tool for cataloguing and researching the production, use and circulation of Fifteenth-century printed woodcuts, in collaboration with the Department of Engineering Science at the University of Oxford (Visual Geometry Group, coordinated by Professor Andrew Zisserman).

The final objective is a system for searching datasets of Fifteenth-century printed images based on the integrated application of both instance-level and category-level image search. Instance-level enables all instances (prints) of a particular woodcut to be matched (retrieved from the dataset). Category-level enables all woodcuts illustrating a category (such as “containing a dog or an XX”) to be retrieved.

The first step of the instance-level process is the application of automatic object retrieval technologies, which seemed particularly suitable for tracking and locating the recurrences of the same woodblock through a potentially endless number of editions by the same or by a different printer, of the same or of a different text. After being uploaded to an online repository, each image, or a selected region of interest within the image itself (in this case, a particular woodblock or part of it), can be used as a query. The object retrieval software will automatically return all the images that contain the query region within seconds (Figs. 1, 2, 3, 4).⁸

Technically speaking, the retrieval system first detects hundreds of points of interest and extracts corresponding visual features from the query image; afterwards, it encodes the image as a single vector considering all its visual components as if they were a “bag” (multiset) of visual words (“bag of words” method). The vector is then compared with the whole dataset and a ranked list is produced, in which those images with the strongest correspondence to the initial query vector appear first. Finally, the initial ranking list is re-ranked according to the geometric consistency between each top dataset image and the query. This instance-level object retrieval pipeline is particularly useful as it allows scholars to know exactly and quickly which images appear and where without having to physically go through dozens of physical volumes or digital reproductions, often not easily accessible. Thanks to its flexibility and reliability, this retrieval

⁵ Forme: “The chase and its contents (type, blocks, etc.), prepared for printing. [...] When using a hand press, each sheet of a book is printed from two formes (one for each side of paper), the inner and the outer formes. The outer includes the first page of the resulting gathering, plus those other pages necessary for the correct imposition of the appropriate format; the inner forme includes the remaining pages of the gathering” [14, vol. II, pp. 730–31].

⁶ Fundamental studies on these aspects include: [3–13].

⁷ In particular, the Broadside Ballads Project (Bodleian Library, Oxford), the online database BSB-Ink (Bayerische Staatsbibliothek, Munich), the Icono15 Project (Bibliothèque Nationale de France, Paris), the Index of Christian Art (Princeton University) and the Arkyves database have been the inspiration behind different trials to apply various kinds of digital reproduction and image-searching technologies to Fifteenth-century printed book illustrations. See them at

Footnote 7 (continued)

<http://www.bodley.ox.ac.uk/ballads/> and at <http://zeus.robots.ox.ac.uk/ballads/>; <http://inkunabeln.digitale-sammlungen.de/start.html>; <http://ica.princeton.edu/range.php>; <http://arkyves.org>.

⁸ For the outline and development of this method, see [1, 2].

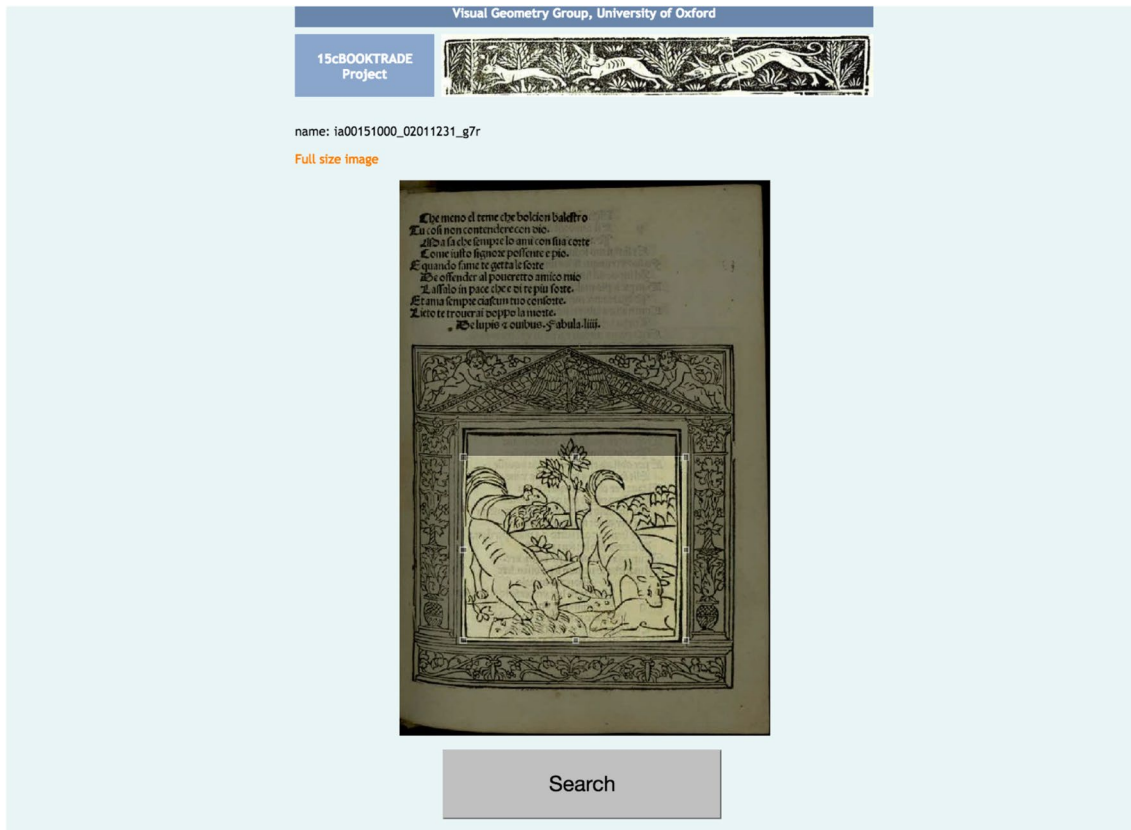


Fig. 1 A screenshot from the 15cBOOKTRADE visual recognition searching demo (© VGG & 15cBOOKTRADE Project). The image is a digital reproduction of leaf g7r of the *Aesopus moralisatus*, Venice: Manfredus de Bonellis, de Monteferrato, 31 Jan. 1491, copy owned

by the Biblioteca Corsiniana (Rome, Italy), shelfmark 51.E.54 (ISTC ia00151000; MEI 02011231). Part of the picture has been selected in order to be compared with the rest of the dataset

system has also been applied by the Visual Geometry Group to many other datasets, such as the British Library’s “1 million images” dataset and the Bodleian Library Ballads dataset; these applications suggested that it might be profitably applied to Fifteenth-century printed book illustrations.⁹

In comparison with other categories of images, such as those found in single sheet ballads or prints, scholars aiming to catalogue and classify Fifteenth-century printed book illustrations have to deal with an additional level of complexity, due to the technical constraints of the printing process of a book, which can run to hundreds of pages and contain several illustrations, occasionally repeated. In order to find out how many and which illustrations were used in a book, it is necessary to map in detail their presence inside the book itself.

Every image in this database is named with a unique identifier which brings together three elements:

- ISTC number,
- MEI number of the copy portrayed in the picture,
- foliation.

From this sequence, it becomes immediately clear in which edition, in which copy and where exactly in the copy the searched image can be found.¹⁰

⁹ For the Bodleian Broadside Ballads database see note 7 (the visual retrieval system is available at <http://zeus.robots.ox.ac.uk/ballads/>). See the British Library “1 million images” demo at <http://www.robots.ox.ac.uk/~vgg/research/BL/>.

¹⁰ The image-retrieval database currently contains around 2500 images and is hosted on the Visual Geometry Group website with restricted access. Over the next months, it will become publicly accessible and direct links between each image and the ISTC and MEI records, respectively related to the edition and to the copy from which the image was taken, will be established, making the whole system even more interoperable and accessible. Up-to-date information on its development and on the web address will be available on the 15cBOOKTRADE Project website dedicated page (<http://15cbooktrade.ox.ac.uk/illustration/>).

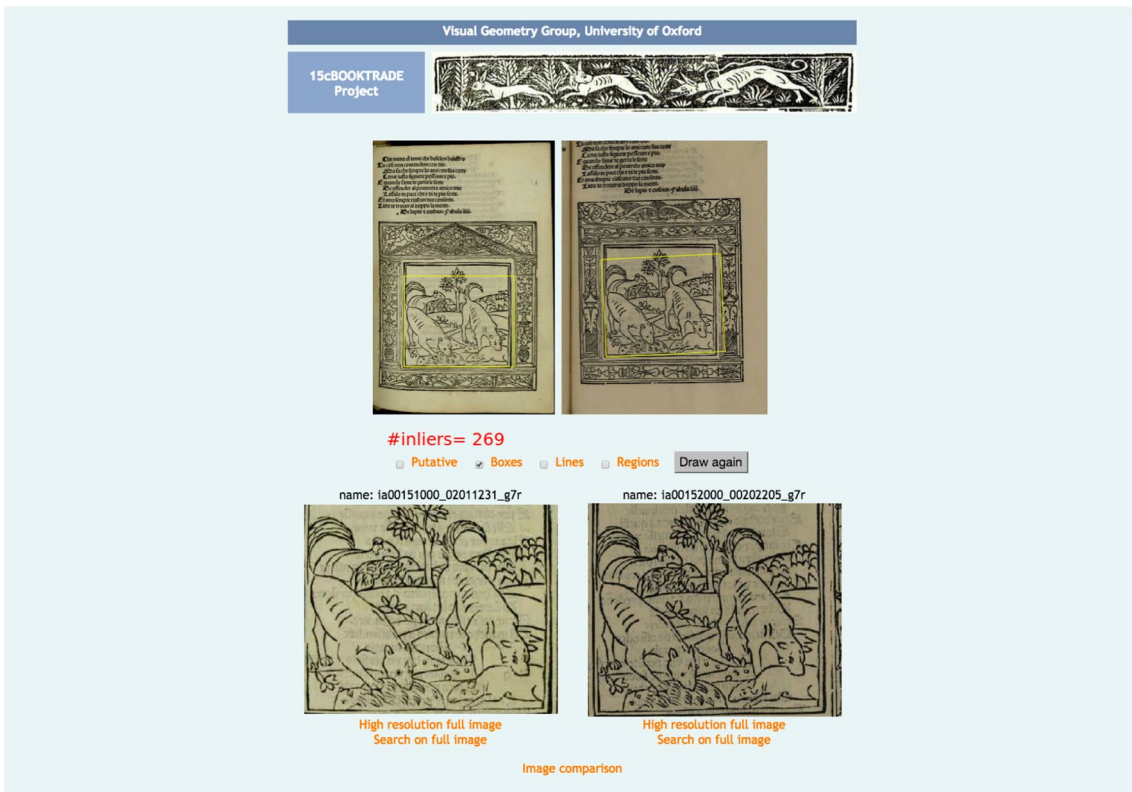


Fig. 2 The result of the query in Fig. 1. The image-matching software detects the recurrence of the query image in one more edition: *Aesopus moralisatus*, Venice: Manfredus de Bonellis, de Monteferrato, 15 Feb. 1491, copy owned by the Fondazione Giorgio Cini (Venice,

Italy), shelfmark FOAN TES 10, leaf g7r (ISTC ia00152000; MEI 00202205). As the reader will notice, in this case the same woodblock was re-used by the same printer within two different frames

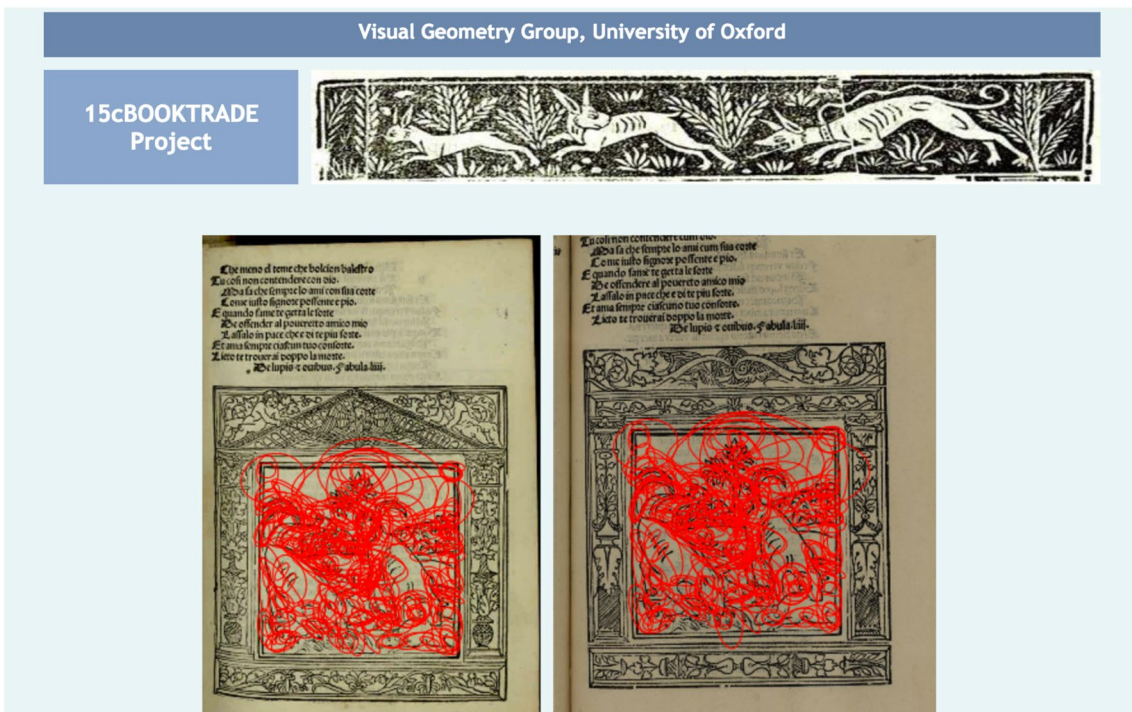


Fig. 3 A detailed comparison between the visual semantic regions of the two images using the “bag of visual words” method



Fig. 4 The results of a different query in the 15cBOOKTRADE image-retrieval demo. The query image stands on the top right side of the screenshot. As the reader will notice, the presence or absence

of colours, as well as the different quality of paper do not affect the effectiveness of the matching system

In particular, the image-matching tool is able to detect the recurrences of certain woodcuts in different editions, which enables us to explore how printed material may have been exchanged between or copied by different printers; it also harvests the recurrence of the same woodcut within a single edition. In this case, by localising precisely every single occurrence of one single woodcut throughout the edition, it becomes evident when it appears more than once within a single printing sheet, although in different combinations.

This suggests that the printer had more than one woodblock of that kind at his disposal. After completing this operation for all the woodcuts, providing the exact number and location of their recurrences throughout the book, we are able to calculate exactly how many blocks were used in one edition and in which combinations.

Reconstructing the composition of the printing sheet becomes useful when trying to analyse the working practice of a printer. While we know a lot about the type case of individual printers, we know next to nothing about their possession of woodblocks. Therefore, this systematic approach is shedding new light on Fifteenth-century printing.

Moreover, when considered not only for their artistic quality but primarily for their content and iconographic features, printed images in early editions have a special value in the reconstruction of the transmission of the text in print. By looking at the iconographic apparatus and its relationship with the text, and by investigating the sources of both

the iconographic and textual tradition, scholars can assess otherwise unknown business relationships between printers, authors and illustrators. They can also explore the development of a certain iconographic cycle or artistic style, school or person. Ultimately, scholars can uncover links with the earlier, as well as with the later, manuscript and in-print transmission.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Bergel, G. et al.: Content-based image recognition on printed broadside ballads: The Bodleian Libraries’ ImageMatch Tool. Paper presented at: IFLA WLIC 2013—Singapore—Future Libraries: Infinite Possibilities, in Session 202 (Art Libraries with Rare Books and Manuscripts). <http://library.ifla.org/209/1/202-bergel-en.pdf>
2. Chung, J.S., et al.: Re-presentations of art collections. Workshop on computer vision for art analysis, ECCV. <https://www.robots.ox.ac.uk/~vgg/publications/2014/Chung14/chung14.pdf> (2014)
3. Donati, L.: Osservazioni sui libri silografici. In: Studi di bibliografia e di storia in onore di Tammaro de Marinis, pp. 207–264. Verona, Stamperia Valdonega (1964)

4. Field, R.S.: Fifteenth century woodcuts and metalcuts from the National Gallery of Art, Washington, DC Publications Dept. National Gallery of Art, Washington, DC (1965)
5. Hind, A.M.: A note on the printing of early woodcuts. *Print Collect Q.* **15**, 131–143 (1928)
6. Hind, A.M.: An introduction to a history of woodcut: With a detailed survey of work done in the fifteenth century. Dover, New York (1963)
7. Kok, C.H.C.M.: Woodcuts in incunabula printed in the low countries. HES & De Graaf Publishers, Houten (2013)
8. Landau, D., Parshall, P.: The renaissance print, pp. 33–38. Yale University Press, New Haven (1994)
9. Lippmann, F.: The art of wood-engraving in Italy in the fifteenth century. Bernard Quaritch, London (1888) (facsimile edition: G. W. Hissink, Amsterdam (1969))
10. Palmer, N.: Blockbooks, woodcut and metalcut single sheets. In: Coates, A. (ed) A catalogue of books printed in the fifteenth century now in the Bodleian Library, pp. 1–50. Oxford University Press, Oxford (2005)
11. Palmer, N.: Woodcuts for reading. The codicology of fifteenth century blockbooks and woodcut cycles. In: Parshall, P. The woodcut in fifteenth century Europe, pp. 93–117. National Gallery of Art, Washington, DC (2009)
12. Parshall, P., Scotch, R. (eds.): Origins of European printmaking: fifteenth century woodcuts and their public, exhibition catalogue curated. National Gallery of Art, Washington, DC (2005)
13. Pollard, A.W.: Italian book-illustrations and early printing: a catalogue of early Italian books in the library of C. W. Dyson Perrins. Oxford University Press, Oxford and Bernard Quaritch, London (1914)
14. Suarez, M.F., Woudhuysen, H.R.: The Oxford companion to the book. Oxford University Press, Oxford (2010)



Matilde Malaspina is a DPhil student at the University of Oxford and a member of the 15cBOOKTRADE Project, coordinated by Dr. Cristina Dondi. She received her BA and MA from the Università Cattolica del Sacro Cuore (Milan, Italy), where she specialised in Medieval and Humanistic Philology. Her doctoral research concerns Fifteenth-century printed book illustrations, with a focus on Italian illustrated editions of Aesopian texts. In 2015–2016 she was awarded a 6-month residential

scholarship at the Fondazione Giorgio Cini (Venice, Italy) to collect and analyse material from the Essling collection of early printed books. She was recently awarded a residential fellowship at the Huntington Library (San Marino, CA) for spring 2018.



Yujie Zhong is a 4th-year DPhil student at University of Oxford. He works on computer vision and machine learning. His research interests mainly focus on content-based image retrieval, including retrieving objects, faces and compound queries.