

## Information discovery using search by image: applications in the humanities

Paul Nieuwenhuysen

Vrije Universiteit Brussel, B-1050 Brussels, Belgium

Paul.Nieuwenhuysen@vub.be



This contribution deals with searching and finding information by using an image as query on the internet and WWW. The overview covers available systems, limitations of their power, and enabling underlying technologies, but the emphasis is on applications in humanities. The audience is motivated and enabled to apply this relatively new method to discover information and to support other potential users.

**Keywords:** Reverse image search, search by image, information discovery, internet, Google

### Introduction

This overview is based on a continuing investigation of the power, applicability, usefulness and limitations of search by image through the Internet / WWW. In this relatively new method for information retrieval, a query does not consist of text but of an image file. The search results lead to images on the WWW and to related documents. The popular Google search system offers this WWW search method as “Reverse image search(ing)”. Other terms used for this method are

- Search(ing) by example
- Reverse image lookup = RIL
- Backwards image search(ing)
- Inside search(ing)
- Content-based information retrieval = CBIR

Furthermore, Google Images even supports a search query that consists of a combination of an image with text.

### Findings

Several online services are available free of charge to search by image.

Differences among these services are substantial.

Google Images can reveal images present on the Internet, which are duplicates / copies of the query / source image. The success is quite variable from case to case.

This performance level concerning recall is strongly correlated with the performance of a more classical Google search by text to find copies of the query/source image file on the Internet.

Reverse image search as offered by Google can even reveal images that are modified versions of the query / source image; more specifically, modified versions can differ from the source image in size and in colors. The system can also reveal a fragment and even a modified / edited fragment of the source image, when this is included in an image present on the internet / WWW.

Three services to reveal copies of an image on the Internet / WWW have been compared. The outcomes have demonstrated that this type a searching can not only reveal

images in simple classical formats such as jpeg / jpg and ping / png, but even copies of the query image that are embedded in a more complicated file such as a PDF or a set of slides. Furthermore, this investigation has led to a ranking for the number of copies found, and to the same ranking for the precision of the search results, as follows:

1. Google, the well know general web search system
2. Yandex, the general web search system that is popular mainly in Russia
3. TinEye, the pioneering system that is dedicated to search by image only

Our tests have demonstrated that since 2014 Google reverse image search can not only find images that are visually similar to the query / source image, but can even retrieve images that are semantically similar / related to the query / source image, while they have no elements in common with the query image. The search results may also include a description of the subject of the image, and this can of course be interesting if the user has not yet much knowledge about the subject, so that using a specific text query becomes possible. Furthermore, other information related to the image and relevant links may also be included in the search results.

The performance of search by image to find images that are semantically similar to the query/source image is improving.

Recently, search by image is applied not only by systems that are primarily search services, but also by systems in which images are important. Examples are systems that offer stock photos (dreamstime.com, gettyimages.com, shutterstock.com) and also Pinterest that allows users to collect or “pin” images in virtual, digital, so-called pin boards. Our comparison of Google with Pinterest has shown that Pinterest has become an additional and even competing system that allows us to start from a selected source image to find related, relevant images; furthermore, any of these found relevant images may lead to information that is related to the selected source image.

Not only pure, simple search either with words or with a source image is possible by the freely available search system that is offered by Google Images, but a search query can also consist of a combination of an image with words. This allows us to combine the strengths of more classical text retrieval with the more recent search by image. Our tests have shown that this allows us to obtain search results with a precision that is higher than when only one of both search methods is used.

The progress described above in automatic analysis of images to determine some of their contents / meaning / semantics is also reflected by the improvements in automatic categorization / classification of images based on their contents. A popular example of such a system that is freely accessible and usable is Google Photos at <https://photos.google.com/>. Our case study has shown that progress is indeed significant and promising.

Trends and recent developments include:

- a. Search by image to find not only related images but even videos (see for instance the company and system Shutterstock that makes available images and videos).
- b. Instead of using an available image file as query, application of a smartphone with inbuilt camera to make a photo and to submit the photo file directly and immediately as a query for a search by image (see for instance Google Lens).

Progress in visual search is supported by:

- a. Refinement of the search system algorithms to detect common elements in images
- b. Improvement of estimating the meaning / contents / semantics of an image from its context on a web page and website

- c. Increasing indexing by search systems of documents in a format that is more complicated than html / xml
- d. Implementation by website creators and exploitation by search systems of a common metadata system (see for instance schema.org)
- e. Progress in artificial intelligence to grasp the meaning of images

## Applications

### Reveal copyright infringements

Starting from an image that you have created or that is affiliated with your organization, you may find copies / duplicates or even modified versions on the WWW. This can reveal copyright infringements.

### Asses the impact of your image(s)

In a more positive way, starting from a source image that you have created or that is affiliated with your organization, you can assess the impact of that images on a worldwide audience. For example: Curators or owners of a collection of objects can assess the impact and reuse of photos of the physical objects in their collection, on a worldwide scale.

### Find a more suitable version of an image

Starting from some image that you have not created, but that you consider as interesting, and that is perhaps not the original version and for which the creator/author is not indicated, you may find other and better versions that are more suitable for your application and need. Also you may find the author(s) on the WWW, which can be useful to obtain more information or to discuss possible copyright linked to the image.

### Reveal misinformation

Also searching by image may allow us to discover that the image that illustrates and supports a document is NOT real / authentic, but that is has been copied from another site, from another context and perhaps that it has even been modified / changed / doctored, to support the text, the claims of the author of the document.

### Find information and other images, all related to the source image

Starting from some interesting source image, you may

- a. identify the contents of that image or reveal information related to that image
- b. find semantically related images; in other words, you may discover images with a subject that is related to the subject of that source image.

More concrete examples of application areas:

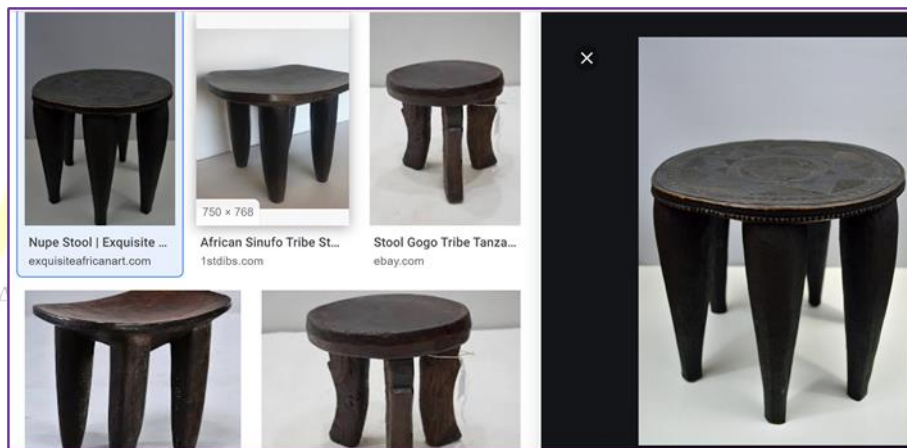
- a. an art object
- b. a building, a monument...
- c. a landscape, a place, a mountain...
- d. a flower, a plant, a plant disease
- e. food (for instance to find information on name, nutritional value...)
- f. the brand / producer / seller of a product
- g. a person (lowering privacy)
- h. a human disease, for instance by using histopathology (see for instance <https://techcrunch.com/2019/07/19/googles-smily-is-reverse-image-search-for-cancer-diagnosis/>)


The following gives a concrete, practical, realistic and recent example, as illustrated in Figure 1. Presume that you want / need more information related to an object (in this example: a wooden seat / stool). Then you can use an existing photo or make a photo of the object, and submit this as a query in a search by image. The results of this search action include small, so-called “thumbnail” images. Then you can select one of these thumbnail images, of course one that looks very similar to your source / query image and activate the

underlying hyperlink. If the link has not been broken recently, then this brings in many cases you to a webpage that shows a larger version of the selected image, plus related information (in this example you can read the description “Nupe stool from Nigeria”). This acquired information allows you to formulate a classical search query with words (in this case ‘nupe stool nigeria’) and to submit this to some WWW search system or to an online encyclopedia (in this example Wikipedia). The search result gives you more detailed information (in this case that Islam religion is important for the Nupe people and that Nupe are famous for their wooden stools with geometrical patterns carved on the surface); the search result can even show a photo of a similar object in a museum collection. So the procedure in this example was simple, fast and productive; in general this way of working is attractive and in many cases productive / satisfactory.



Visually similar images



	<p><b>Tribe:</b> Nupe</p> <p><b>Origin:</b> Nigeria</p> <p><b>Approx Age:</b> Early – Mid 20th Century</p> <p><b>Materials:</b> Wood</p> <p><b>Dimensions cm:</b> 36 (tall) x 40 (wide)</p> <p><b>Ref. Number:</b> 1011</p>
	<p><b>£800.00</b></p> <p><b>Description:</b> An outstanding Nupe Stool from Nigeria. Showing a beautifully carved seat of circular form with a geometric design which is typical of the Nupe stool. Being carved out of one piece of wood, the numerous legs that support the seat make it a wonderfully stable stool. This has</p>

**WIKIPEDIA**  
The Free Encyclopedia

Main page  
Contents  
Featured content  
Current events  
Random article  
Donate to Wikipedia  
Wikipedia store

Interaction

Help  
About Wikipedia  
Community portal  
Recent changes  
Contact page

Tools

What links here  
Related changes  
Upload file  
Special pages

## Nupe people

From Wikipedia, the free encyclopedia

*This article is about the ethnic group. For fraternity members called nupes, see [Kappa Alpha Psi](#). For the British trade union, see [National Union of Public Employees](#).*

? This article includes a list of references, but **its sources remain unclear** because it has **insufficient inline citations**. Please help to **improve** this article by **introducing** more precise citations. *(April 2019)* [\(Learn how and when to remove this template message\)](#)

The **Nupe**, traditionally called the **Tapa** by the neighbouring Yoruba, are an ethnic group located primarily in the Middle Belt and northern **Nigeria**, and are the dominant group in **Niger State**, an important minority in **Kwara State** and present in Kogi State as well.

**Contents** [hide]

- 1 History
- 2 Population and demography
- 3 Traditions, art and culture
  - 3.1 Examples of Nupe art
- 4 References
- 5 Notes

Many Nupe were converted to [Islam](#) at the end of the eighteenth century by Mallam Dendo, a wandering preacher, and were incorporated into the [Fulani Empire](#) established by the [Jihad](#) led by [Usman dan Fodio](#) after 1806.

out in certain areas. Their art is often abstract. They are well known for their [wooden](#) stools with patterns carved onto the surface.



Wooden oval stool with incised carving; Los Angeles County Museum of Art.

Figure 1. Example of the steps executed to discover information that is related to a particular object, starting with a search by image.

### **Increase the precision of search results by adding text to the source image**

Furthermore, including some text in the query besides the image may increase the precision of the results, even when not enough knowledge is available in advance, so that only one or a few unspecific search words can be used.

### **Increase the precision of search results by adding an image to a text query**

Consider the scenario in which you have already sufficient information/knowledge in advance to formulate and submit a classical, textual, specific, focused search query; even then, including an image to the text query may be useful to increase the precision of the search results.

### **Increase the usability of a digital library of images**

Automatic classification of images may be useful for managers of a digital library that includes images, to increase the quality / usability of their system by increasing the browsing and retrieval efficiency.

### **Recommendations for practitioners**

The growing success of the search methods that include an image in the query to find relevant information leads us to a few recommendations:

- a. To find relevant information, these recent, additional search methods should be considered besides more classical methods, by librarians and information intermediaries in general, and also by end-users of information discovery systems.
- b. As a consequence, search by image deserves a place in educational courses and tutorials on information and media literacy.
- c. Authors and publishers in general want to create their publications and make these available in such a way that they rank high in the results of relevant search and discovery systems. Therefore it is good practice to take into account the workings of at least the classical, popular, search services, in the creation and optimization of their website(s). Not only the texts in a website should be considered, but also images, to optimize

--for a relatively classical search with a text query to find images,

--for a more recent search by image, or

--for a search with a query that consists of text plus an image file.

More concretely, website developers should try to publish their meaningful images in such a way that these can be well harvested, analyzed and included in the database index of relevant search systems in an effective way.

### **Conclusions**

Information discovery is enhanced by recent methods that involve images:

- a. Search by image is evolving to a powerful, additional method to tackle information needs that are difficult to handle with more classical methods.
- b. Information discovery is assisted by automatic classification of images and by recommendation services based on image similarities.

Furthermore, using a combination of text with an image in a search query can increase the precision of the search results, in comparison with a more classical pure text search or with a pure search by image.

### **References**

- Kelly, E. (2019) Reuse of Wikimedia Commons Cultural Heritage Images on the Wider Web, Evidence Based Library and Information Practice, 14.3.

- Marques, O. (2016) Visual Information Retrieval: The State of the Art. *IT Professional*, 18, Issue: 4, DOI: 10.1109/MITP.2016.70.
- Nieuwenhuysen, P. (2013) Search by Image through the WWW: an Additional Tool for Information Retrieval. Full text published in Proceedings of the international conference on Asia-Pacific Library and Information Education and Practices = A-LIEP 2013 “Issues and challenges of the information professions in the digital age” held at Pullman Khon Kaen Raja Orchid Hotel, in Khon Kaen City, Isan, Thailand, 10-12 July 2013 [online] <http://aliep2013.com/index.php/table-of-contents> <http://aliep2013.com/images/download/pdfs1/PaperNo41.pdf> & full text available also online free of charge from [https://www.researchgate.net/profile/Paul\\_Nieuwenhuysen/contributions](https://www.researchgate.net/profile/Paul_Nieuwenhuysen/contributions)
- Nieuwenhuysen, P. (2014) Search by image through the Internet: applications and limitations. In *Libraries in the Transition Era: New Space – New Services – New Experience*. The Proceedings of the Seventh Shanghai International Library Forum, organized by the Shanghai Library, in Shanghai Library, China, 9-11 July 2014 <http://www.libnet.sh.cn/silf2014/english/index.htm> Shanghai: Shanghai Scientific and Technological Literature Press, <http://www.sstlp.com>, 476 pp, ISBN 978-7-5439-6289-7. pp. 145-155.
- Nieuwenhuysen, P. (2015) Search by image through the Internet: an additional method to find information. In *Transforming Libraries and Librarianship*, Delhi: KBD Publication, 502 pp. Edited by Sanjay Kataria, John Paul Anbu, Shri Ram, Nirmal Kumar Swain, Naresh Singh Bhandari, 2015, ISBN: 978-81-907999-6-6, pp. 179-194.
- Nieuwenhuysen, P. (2016) Information discovery on the Internet, using a search query that consists of text & an image. in A-LIEP 2016, proceedings of the 7th Asia-pacific Conference on Library and Information Education and Practice, in University of Nanjing, Nanjing = Nanking, Jiangsu province, China, Edited by Jianjun Sun, Qinghua Zhu, Christopher Khoo Shiyan Ou.. <http://aliep2016.nju.edu.cn>, PDF file, pp. 99-112. Available free of charge from <http://aliep2016.nju.edu.cn/files/A-LIEP%202016%20Proceedings.pdf>
- Nieuwenhuysen, P. (2017) Information discovery: A picture is worth a thousand words? In *Proceedings of the International Conference on the Future of Libraries (ICFL), From Promises to Practice*, Editor-in-chief M. Krishna Murthy, Editors B. Ramesha, B. Subhash Reddy, organized by DRTC = Documentation Research and Training Centre, Indian Statistical Institute, Bangalore @ ISI = Indian Statistical Institute in Bangalore = Bengaluru, Karnataka, India, 15-17 November 2017, Published by DRTC, Indian Statistical Institute, 2017, 602 pp., ISBN 978-81-92179-79-7, pp. 313-324 & available from [https://www.researchgate.net/profile/Paul\\_Nieuwenhuysen/contributions](https://www.researchgate.net/profile/Paul_Nieuwenhuysen/contributions)
- Nieuwenhuysen, P. (2018a) Information discovery based on the emerging technology to analyse digital images. In *Reshaping libraries: emerging global technologies and trends*. ICRL-2018. Papers of the First International Conference on Reshaping Libraries: Emerging Global Technologies and Trends, organized jointly by DELNET-Developing Library Network, Ambedkar University Delhi in Association with the Society for Library Professionals (SLP) and Special Libraries Association (SLA, USA) Asian Chapter, February 1-3 at Jaipur, India. Chief Editor: H.K. Kaul, Editors P.K. Jain, Debal C. Kar, Sangeeta Kaul. Published by DELNET, New Delhi. ISBN 978-93-82735-13-7. 230 pp., pp. 1-14 & available from [https://www.researchgate.net/profile/Paul\\_Nieuwenhuysen/contributions](https://www.researchgate.net/profile/Paul_Nieuwenhuysen/contributions)



- Nieuwenhuysen, P. (2018b) Information Discovery and Images A Case Study of Google Photos. in: 5th International Symposium on Emerging Trends and Technologies in Libraries and Information Services (ETTLIS), in Noida, India, 21-23 Feb. 2018, INSPEC Accession Number: 18150464, DOI: 10.1109/ETTLIS.2018.8485238, Publisher: IEEE <https://ieeexplore.ieee.org/document/8485238> & available from [https://www.researchgate.net/profile/Paul\\_Nieuwenhuysen/contributions](https://www.researchgate.net/profile/Paul_Nieuwenhuysen/contributions)
- Nieuwenhuysen, P. (2018c) Information discovery using search by image: a comparison of Google and Pinterest. In The Proceedings of the 9th Shanghai International Library Forum = SILF 2018, Library for all: towards a smarter and inclusive society, October 17-19, 2018, at the Shanghai Library, in Shanghai, China, 464 p., pp. 400-412, ISBN 978-7-5439-7748-8, <http://www.sstlp.com> & available from [https://www.researchgate.net/profile/Paul\\_Nieuwenhuysen/contributions](https://www.researchgate.net/profile/Paul_Nieuwenhuysen/contributions)
- Nieuwenhuysen, P. (2019) Finding copies of an image: a comparison of reverse image search systems on the WWW. In Collaboration – Impact on Productivity and Innovation : Proceedings of 14th International Conference on Webometrics, Informetrics and Scientometrics & 19th COLLNET Meeting 2018, December 5-8, 2018, University of Macau, Macau, Edited by Markscheffel, Bernd; Kretschmer, Hildrum. 130 pp. Available online, free of charge from [https://www.db-thueringen.de/receive/dbt\\_mods\\_00039355](https://www.db-thueringen.de/receive/dbt_mods_00039355), DOI: 10.22032/dbt.39355. 10 pp.
- Thompson, S., & Reilly, M. (2017) “A picture is worth a thousand words”: Reverse image lookup and digital library assessment. *Journal of the Association for Information Science and Technology*, 68, pp. 2264–2266. doi:10.1002/asi.23847

### **Biography**

Paul Nieuwenhuysen is professor emeritus @ Vrije Universiteit Brussel. His functions have included: member of the management board of the University Library, librarian for science and engineering, and teaching on online retrieval and presentation of scientific information. In the inter-university postgraduate program in Information and library science @ University of Antwerp, he has been guest professor.

At the University of Antwerp he received the degrees of Licentiaat Physics in 1974, Doctor in Science in 1979, the Belgian post-doctoral degree in 1983, and the inter-university postgraduate degree in Documentation and library science in 1986.

He has organized 15 international training programs on management of information in science and technology in Brussels; afterwards he has co-organized international training programs on information technology for scientific information management at University of Antwerp.

More information is available from <http://homepages.vub.ac.be/~pnieuwen/>