Application of Digital Images and Corresponding Image Retrieval Paradigm

Marina Ivasic-Kos
Faculty of Informatics and Digital Technologies, University of Rijeka, Croatia

Abstract

We live in a world where digital images are constantly generated during our daily activities, whether private or business. They play an important role in our private life, showing important moments, people, places, or events and keeping their memory. Images are unavoidable in business, especially in digital marketing, web sales, social networks, medicine, security, and education. In general, images contribute to a better understanding of the message, increase the attractiveness of textual content, provide a better user experience, and can convey emotion quickly. The key advantage of the image is that very often, even a cursory glance at the image is enough to convey a message and arouse emotion and interest. But with the increase in digital image numbers, storage, organization, and retrieval problems arise. The paper describes the importance of images in different areas of application and different image retrieval paradigms that include text-based, content-based, and combined approaches. Also, the most popular image search tools and cloud storage services are compared and discussed. The conclusion comments on the applicability of existing approaches to image searches in different application domains and highlights the advantages and disadvantages of each of the approaches.

Keywords: digital images; content-based image retrieval; text-based image retrieval; semantic search; cloud storage services
JEL classification: O14
Acknowledgments: Croatian Science Foundation supported this research under the project IP-2016-06-8345, "Automatic recognition of actions and activities in multimedia content from the sports domain" (RAASS), and by the University of Rijeka (project number 18-222).

Paper type: Research article
Received: 12 Jun 2022
Accepted: 4 Jul 2022

DOI: 10.54820/entrenova-2022-0030

Introduction

An important feature of images is that they can easily attract our attention. Namely, humans are visual beings, and a large part of the human brain is dedicated to processing and analyzing visual information. In a very short time, people can learn the meaning of an image, recognize familiar objects and people, and analyze the entire scene. Some concepts can be better explained in pictures than in text, and information can be obtained more simply in less time than needed to read a long sentence or describe an image. That is why images have always played an important role in a person's life. Today we can hardly imagine an ordinary day without using images to document our day-to-day lives and our most precious moments in life's private or business sphere. The development of technology, especially the ubiquity and availability of mobile phones and camera devices in parallel with the increasing quality of digital photos and computer speed and power, has encouraged mass image generation and use in various applications.

We take images at birthdays, parties, trips, concerts, sports, and other private events (Figure 1). Images are a way to capture an important moment in our lives, such as the birth of a child, first steps, promotions, a wedding, to evoke the emotion we want to relive, to capture memories of important people, places, or events and to preserve them from forgets. Images can refresh memories and emotions, convey experience and information about the moment in which they were created, and serve as a confirmation of authenticity and the perfect medium for looking into the past. Images are also important as a means of communication and an easy way to convey a message, and share memories, experiences, and emotions with our loved ones, family, and friends.

Figure 1
Various examples of private images that represent visited places and important experiences but also preserve the memory of the past

Source: Googe search

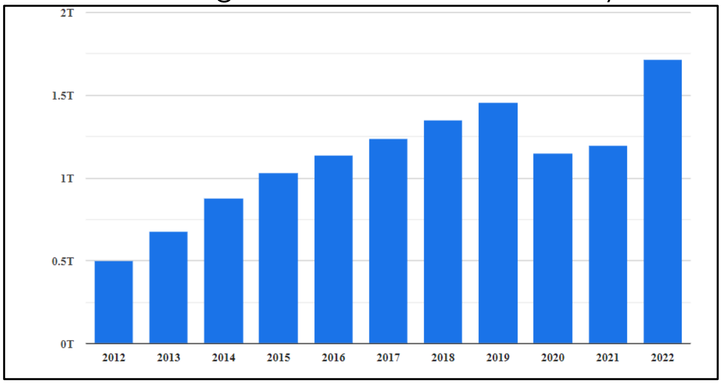
Images are becoming necessary in several business areas, such as economics and digital marketing, tourism, medicine, security, sports, and entertainment (Figure 2). This ubiquitous increase in the number of images related to different application domains has been further influenced by improving camera characteristics and image quality, increasing computer speed and power, increasing capacity and reducing the cost of story space, and developing communication infrastructure for their distribution. Also, digital images are easily stored as image data files (e.g., JPG, BMP, TIF, GIF, PNG), easily and quickly distributed, and modified if necessary with appropriate software tools. Furthermore, digital images do not lose quality over time, but they are as safe as their storage medium.

Figure 2
Examples of images from different application domains such as e-sales of goods and real estate, medicine, security, sports, autonomous vehicles, sports

Source: Consido Official, Pynck Consido Official, Radiopedia Official, Googe search

Today, private and business collections are growing rapidly, and Mylio.com estimates that the total number of images taken per year is over 1.4 T or 3 billion images per day. Photos statistics estimate that 1.72 trillion images will be taken worldwide in 2022, which is 54,400 per second or 4.7 billion per day. The number of images has been linearly growing since 2012 (Figure 3) by an average of 10-14%, so it is expected by 2030 to be twice as many images as now, meaning that 28.6 trillion images will be taken annually. The exception to the expected growth in the number of images in the world was the pandemic years when a decline of about 20 percent was recorded.

Figure 3
Number of images taken worldwide each year from 2012 to 2022



Source: Photos-statistics, 2022

A problem with the growing number of images is the increased space required for their storage. Therefore, they are usually stored on additional electronic media (external disk, memory card, etc.), cloud spaces (Google Photos, OneDrive, Flickr), or stock photo libraries (Flickr, Picasa, Photobucket, Bigfoto, etc.).

Another problem is the search, retrieval, organization, and use of images. It is thought that we could easily retrieve and organize images when it would be possible to automatically describe them with words that are intuitively used when searching for images (Hare, 2006). However, images are generally difficult to classify, and sometimes it is not easy to describe images in words that will meet different requirements and needs of users, so different approaches to image search and retrieval have been developed. The second section describes the different areas of use of digital images, and the third describes recent trends in storing huge image datasets. The fourth section briefly presents the paradigms of image retrieval based on text and image content and combinations of these fundamental approaches. A description and comparison of the most popular image search engines are given in the fifth section. In conclusion, recommendations regarding image search engines are given for image retrieval and search in different application domains, highlighting the advantages and disadvantages of each approach.

Use of digital images in various domains

Searching, sharing, and comparing digital images have been incorporated into various aspects of people's lives. Today, digital images are indispensable in various business applications and domains such as e-commerce, marketing, tourism, publishing, medicine, criminology, sports, culture, entertainment, and others.

Digital marketing and social networks

In economics, especially in e-commerce and digital marketing, images are used for advertising and displaying the properties of products or real estate, inspiring customers and motivating them to purchase (B. Bowers, 2022).

The use of images has proven to be a natural method for increasing the attractiveness of textual content and fluidly presenting content. In addition, images break up large blocks of text and contribute to the easier reading of text, faster memorization and perception of content, and faster searching and finding of information. These facts are actively used on social networks to convey as much information as possible in a concentrated form and attract more audience attention, elicit emotions and achieve greater audience engagement or interaction, better user experience, and a better understanding of the message. E.g., analyses of visits to social networks at BioMed Central (BioMed, 2022) showed that a post on social networks accompanied by an image is ten times more likely to be noticed. Similar experiences are on Facebook, where posts with a picture, for example, get twice as many comments as those without images. Images are also used on Twitter to convey information that cannot be explained within 140 characters of text as allowed. The rest of the social networks look very similar; 63% of the content consists of images (Skimlinks, 2022) that attract the audience's attention and gain interaction by commenting liking, or sharing.

Tourism and culture

Today, the use of images to promote tourist destinations, hotels, restaurants, museums, shops, and various facilities is becoming increasingly popular. In addition to images, short video clips and virtual walks are used, which combine multiple images to cover 360° of horizontal and 180° of vertical field of view to create the impression of 3D space. Some examples of such virtual walks to promote tourist destinations can be (The virtual tour, 2022) website, and (Google Art, 2022), and (EU Virtual Museum, 2022) which are examples of virtual museums.

Medicine

In medicine and dentistry, images are primarily used for diagnostic purposes and collected using various specialized devices such as Colour-Doppler, MRI, and CT. For example, in medicine, a doctor can diagnose diseases more quickly and accurately and monitor a patient's condition by comparing digital images stored on a patient's map. Also, by exchanging digital medical images, a doctor can very quickly get a second opinion from an appropriate specialist from a remote medical center, which benefits the patient and the doctor's professional development.

Surveillance and security

In the surveillance and security of properties and people, the most common cameras are those that shoot 24/7 from a bird's eye view and various devices to capture images in different weather conditions with more accurate measurements, such as radar, LIDAR, and IR cameras. In addition, various scanners that compare a person's biometric features, such as a person's face, fingerprint, palm, or signature, with an image stored in a database are used to authenticate a person in banks or when entering protected facilities. Similarly, vehicle registration plates are used to enter or exit the garage. Storing, exchanging, and comparing digital images of fingerprints, palms, or facial profiles facilitates criminal cooperation with each other and speeds up the detection of suspects.

Image storage spaces and image datasets

Digital images are easy to store and easily and quickly distributed, but the challenge is how to safely store all those images and save them from potential loss. Also, the problem that may arise is the durability of the media on which the images are stored.

Images do not lose quality over time, but they are as safe as the medium on which they are stored. Over the years, the media can become unreadable or outdated and lose data due to inadequate archiving or data protection or because support for the media is no longer available. The solution is a secure platform with a capacity that can increase storage along with an increasing number of images and is convenient. Hard drives and flash drives have a hard time meeting these requirements, so currently, the best solution is to use cloud storage. Some popular image storage in the cloud is Google Photos, IDrive, OneDrive, Flickr, and Dropbox. Of these storages, the largest free space is provided by Google Photos, 15GB, while Flickr and Dropbox provide unlimited space but not for free. All but Dropbox allow accessing files from an unlimited number of devices, which is especially important for business users.
Important features are the ease of use and the ability to upload images from mobile phones since most images, 92.5% (Photos-statistics, 2022), are taken with a mobile phone.

Images can be published on public image libraries such as Flickr, Picasa, Photobucket, Bigfoto, Shutterstock, and others.

However, besides private storage and public libraries, there is a growing need for publicly available datasets to be used for education, research, and other no commercial purposes. These are mainly large sets of images used to train machine learning models for different computer vision tasks in different areas of application that allow automatic image analysis, object detection, or classification. The most well-known datasets are COCO (Lin et al., 2014) and ImageNet (Deng et al., 2009), which contain many tagged images of objects in the natural environment. For example, the COCO datasets currently contain about 330,000 images, of which over 200,000 are tagged with at least one object label from 80 categories. Furthermore, in addition to general-purpose datasets, there are domain-specific image datasets such as the image datasets for the living world - iNaturalist, (Van Horn et al., 2018), the medical image database of the Centres for Disease Control and Prevention (PHIL, 2022), a public fingerprint database (NIST, 2022) or a dataset for the detection of persons in thermal images - UNIRI-ITD, (Kristo et al., 2020).

Image retrieval paradigms

Due to the increase in the number of digital images, ambitious efforts have been made over the past decades to automatically index, classify, and describe different images to simplify their organization, retrieval, search, and use (Datta et al., 2008).

Active research into the problem of image retrieval began as early as 1970 and was marked by two approaches (Rui et al., 1999):

- text-based image retrieval
- content-based image retrieval - CBIR.

Recently, these approaches have been combined, and such retrieval is most often called a semantic-based retrieval, especially if terms at different levels of abstraction are used (Ivasic-Kos et al., 2009). In addition, additional information such as geolocation and hashtags are often included in the search based on image information. Some subtypes of image content search are querying by visual sketch, querying by multiple example images, navigating customized/ hierarchical categories, searching by selected image region, querying by the specification of image features, and multimodal queries that include touch and voice.

Text-based retrieval

When retrieving or searching for images based on text, images are described by keywords and comments (caption, description, image title), which are taken into account when searching. The development of this approach began in the database management community, and images were retrieved by SQL queries based on textual data. The positive characteristics of this approach are the speed and ease of searching, i.e., the successful use of existing technologies, and results close to user expectations. However, to be able to retrieve images, they needed to be described in words, which required a lot of intensive manual work. Therefore, the names of image files or words found in the caption or the comment next to the image were used for image retrieval at the beginning. However, this resulted in images from completely different classes without finding images marked with synonyms or a name in another language (Figure 4).

Figure 4

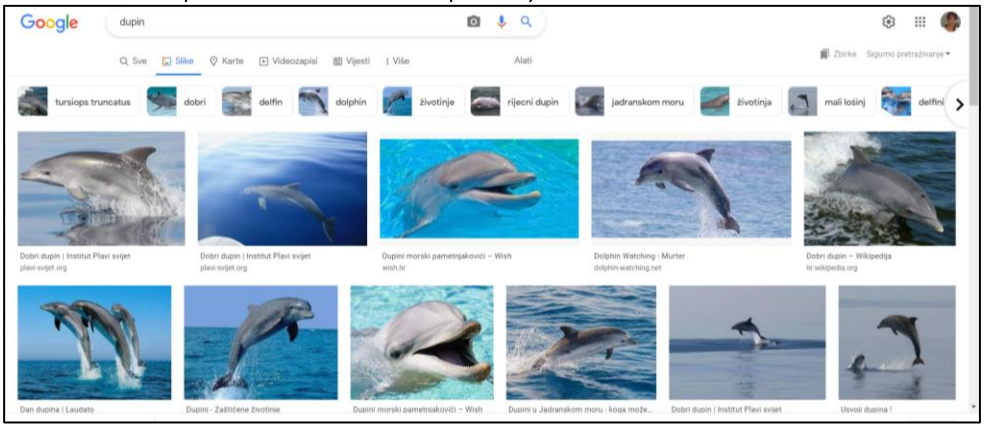
Images as a result of the query "dupin" based only on the text next to the image

Dupin u skoku	**dupin**.jpg	Tehnika **dupin** je	Charles **Dupin**

Source: Googe search in 2012

In recent years, sophisticated algorithms such as visual attention and tensor networks have been intensively developed to better extract contextual information from images (Stefanini, 2021, Wang et al., 2022) so more context-related images to a query term are retrieved. Also, links are offered to images related to synonyms, superclasses, English terms, and terms that often appear together with the query. For example, text query "dupin" in Figure 5, the images related to query class dauphin are retrieved, but also images related to a particular species of dolphin (bottlenose dolphin), superclasses (animals), synonyms (delfin), and words that are often mentioned with the word dolphin in Croatian webpages (Mali Lošinj, Jadransko more).

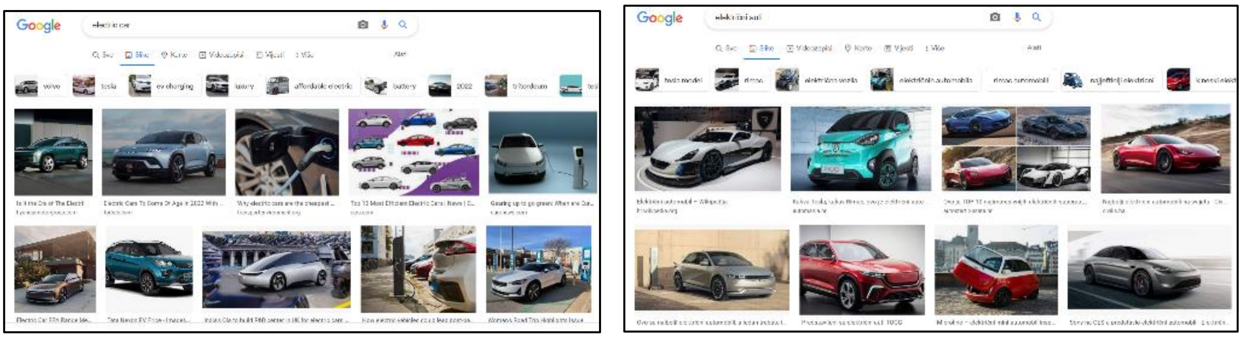
Figure 5
Images as a result of the query "dupin" with access to images related to synonyms, sub-terms, super terms, and frequently related terms



Source: Googe search in 2022

Text-based search depends on the text in which images are described, so the retrieval result depends on the language used to post the query since images are searched on localized pages. An example is searching results in Figure 6 for electric cars using query terms in English and Croatian. Related terms also depend on the language used, so in the case of English, a Tesla car is provided, while Rimac's name is mentioned in Croatian.

Figure 6
Sensitivity of image retrieval on the language used for posting the text query, English (left) and Croatian (right). Different results and related terms are provided.



Source: Googe search in 2022

Content-based image retrieval

Content-based image retrieval (CBIR) began exploring the computer vision community in the early '90s. Intuitively, the main task of the CBIR system is to reduce a large set of images, using the visual content of the image, to a small subset of those with the desired features. The term CBIR was first used by T. Kato in 1992 to describe an experiment to automatically search images from a database based on colour and shape representations (Rui et al., 1999).

A query based on image content is most often defined (Siggelkow, 2002) by one or more image examples (QBE - query by example), a rough sketch or approximation of the shape or colour of the desired object ignoring the background (QBS - query by sketch) and selecting the values of low-level features that the requested image should have (e.g., 70% blue, 30% red).

The most common approach to searching for images by image content is for the user to submit a query image and for QBE to find images identical or similar to the images from the query, i.e., images of the same class (Hare, 2006). The first commercial QBE system, IBM's QBIC, in 1996 considered colour, colour layout, texture, shape, size, orientation, and position of image objects and regions.

Because an image consists of non-meaningful image elements (pixels), the first step in finding or searching for images by content is determining the visual image features used in the search. The visual content of an image is represented with low-level features such as colour, shape, texture, or any other information that can be derived from the image itself. Earlier, various techniques were used, such as colour histograms, edge, interesting point detectors, gradients, and the like, while today, the features are obtained automatically using deep neural networks (Ivasic-Kos et al., 2015, Hrga et al., 2019).

Systems that search for the target image search for whether it exists in the dataset, while those that search by class retrieve images that belong to the requested class. When searching by class, images similar to the target or query images are searched, so similarity metrics must be determined to compare images. Target images are usually sought in forensics (fingerprints, face profiles) or security systems (fingerprints, iris). Search by categories is represented, for example, in medical diagnostics (fractures, lung diseases, etc.) or general-purpose images (city, sea, sports, etc.).

Given the significant visual differences that instances of the same class may have in colour, topology, size, shape, etc., and the similarity between instances of different classes, good image representation plays a crucial role in image classification and comparing images. Significant improvements in image search and retrieval have been driven by the development of deep learning models that have been learned from a large number of images and achieve relevant classification results in the general-purpose image and can distinguish objects from the background and detect objects in images (Buric et al., 2018).

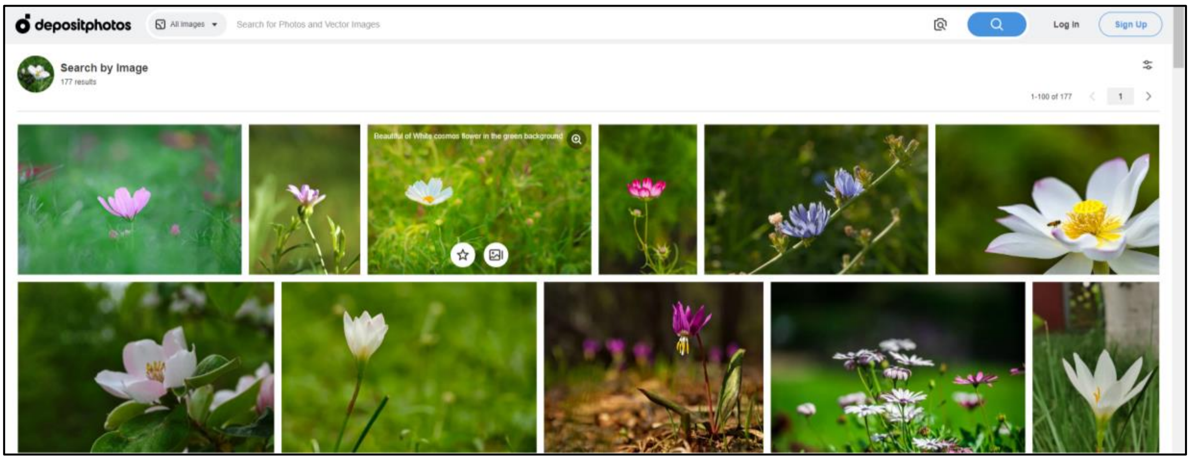
This search approach includes multi-dimensional indexing and data modelling techniques. It is used by some search engines, such as Google Image Search, which indexes images by text and metadata on the web page where the image is located. Also, commercial catalogs maintained by the Getty Information Institute have the same search approach (Siggelkow, 2002) and contain approximately 120,000 words for commenting on objects, images, textures, architecture, and the like.

Figure 7 presents an image query and search results based on visual content and default similarity metrics. Retrieved images are selected based on colour histograms so that the result includes images of different classes (flower, lamb, cat). At the same time, Figure 8 shows retrieval results, including images of the same class with similar colour layouts and visual characteristics.

Figure 7
Query images and retrieved images based on the colour histogram that corresponds to different classes

Source: Google search in 2012

Figure 8
Image query and retrieve images based on colour layout corresponding to the same class



Source: Depositphotos in 2022

Multimodal image retrieval

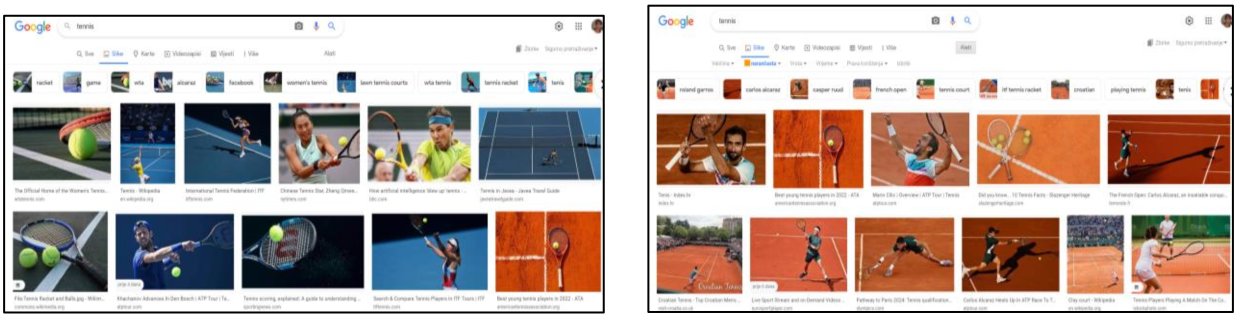
Multimodal image retrieval uses both keywords and visual image properties. Therefore, it is a compromise solution that allows image retrieval according to the semantics included in the text query but follows the required visual characteristics.

Expected results are images that match a given keyword and have the appropriate visual characteristics; most often, it is the chosen dominant colour, size, or type of image. The system first retrieves all images that match the query keyword and then extracts the low-level visual features and gives those images with the most similar features.

Figure 9 shows the search result for the text query tennis. With the additional condition of the dominant colour, e.g., orange, a modification of the result related to tennis played on the ground is obtained.

Source

Figure 9
Image query for text "tennis" (left) and additional colour feature orange (right)



Source: Google search

Query images can be used for the automatic extraction of visual features and, combined with text query (Figure 10), can be used to retrieve images that correspond to text query but are visually similar to query image (Pobar & Ivasic-Kos, 2015).

Figure 10
Text query + image for automatic extraction of visual features

Source: Google search

Comparison of search results of popular image search tools

An image search engine helps you find images on the Internet using keywords, phrases, or example images by ranking images according to relevance.

Today's most common search engines offer image searches like Google, Yahoo, or Bing!, but there are also those specializing only in image searches like Getty, Flickr, Shutterstock, and Yandex. Some engines, such as Getty or Flickr, support image search only by text query, some, such as Yandex or Pexels, only by image example, and some, such as Google Image Search, Depositphotos, and Shutterstock, are both types of search.

Most image search engines work like ordinary search engines by indexing image metadata and storing it in a large database. When a search query is performed, the image search engine searches the index, matching queries to the stored information. The result is a set of thumbnails displayed by relevance. The usefulness of an image search engine depends on the relevance of the results it gives, and the algorithms most often differ in the metrics used to assess the similarity between images and the relevance of the related terms to rank the results (Ivasic-Kos et al., 2014).

Table 1
Comparison of properties of image search engines

Image search tool	Text-based	Content-based	Related terms/crop area	Colour features	Image type	Image Properties	People	Search By	Language
Google	x	x	x	x	x				x
Yahoo!	x		x	x	x		x		x
Big	x		x	x	x				x
Getty	x		x	x	x	x	x	x	
Flickr	x			x	x	x	x	x	
Shutter stock	x	x		x	x	x	x		
Pexels		x	x						N/R
Depositp hotos	x	x	x	x	x	x	x		N/R
Yandex		x	x	x					N/R

Source: Authors'.

The number of available public systems that include image search algorithms that use low-level visual features has been growing in the last twenty years (Wang et al., 2001). For example, some search engines have additional options to recognize the dominant colour in an image based on colour histogram, texture, or geometric image topology. In contrast, others use more sophisticated algorithms to detect people in an image, even their gender, age, position, and the like. These low-level features can be used independently, as in retrieving images based on content or increasing metadata in image search. A comparison of image search engines concerning the search paradigm they use and the low-level features they have involved in search refining is shown in Table 1.

Text-based image search engines like Google, Bing, and Yahoo! offer links to images described by related terms and terms in English or locally and refine search results by including image colour, size and resolution features, and usage rights.

Specialized image search engines such as Getty and Flickr have additional options for searching images by image orientation and camera position. At the same time, Flickr offers search by texture, scene topology, and depth of field. Flickr and Getty also provide a search based on the presence of people on the scene and a refined search by gender, number of people, and age. Getty also allows searching by composition and ethnic data. Such additional search conditions that include automatic detection of persons and classification by sex and age became possible in the era of deep neural networks that today achieve results of classification and detection at the human level (Sambolek & Ivasic-Kos, 2021).

Both Getty and Flicker can define how to choose the result, whether it's the best match, the latest post, or the most popular image. When searching for images, Depositphotos also offers the ability to select images according to the camera's position, such as a bird's eye view, aerial images, and images that include a panorama, portrait, or full body. Yandex and Pexels search engines search images as the best match to an example image but also enable the selection of a part of the image to be searched.

It is also important to note that although images are available on the Internet, each search engine retrieves different sets of images for the same query. It is sometimes more useful to use different search engines to find what we want or need.

Conclusion

The field of automatic analysis of digital images and the development of tools for their use today attracts great interest from researchers. Images play an important role in all spheres of our private and business life. The development of technology, especially the ubiquity and availability of mobile phones and cameras, along with increasing image quality and the speed and power of computers, has encouraged mass image generation and use in various domains, from security, medicine, marketing, and e-commerce to sports and entertainment.

Digital images are easy to store and easily distributed and modified with appropriate software tools. Still, due to the huge amount of image material, the usual storage media is not enough, so it is usually stored in the cloud (Google Photos, OneDrive, Flickr) or photo libraries (Flickr, Picasa, Photobucket, Bigfoto, etc.). Along with many images, there is the problem of retrieving and searching images because there is still no algorithm that would automatically extract image semantics for different application domains, different types of users, and different levels of interpretation. Therefore, two dominant approaches for image retrieval are the content-based and text-based approaches. Image search based on content is natural for some application domains such as criminology, medicine, or biology, where a comparison of images with existing ones in the database is required. However, when we don't even have an image available because we're just looking for it for general-purpose images, it's more natural and intuitive to search based on the text that users would otherwise use to tag or describe those images.

Some search engines, such as Google Image Search, Depositphotos, and Shutterstock, allow you to use both approaches. It is also the practice that low-level features such as dominant colour, scene topology, or texture are increasingly used to filter results. With tools specializing in image searches, such as Getty and Flickr, the search can also be refined according to the image orientation, camera position, and information about people on the scene, such as gender, age, number of people, and the like.

In recent years, the emergence of deep neural networks has significantly improved the results of computer vision methods that can automatically classify an image and detect various objects in the scene, so it seems a natural way to include such information in an increasing number of image search tools.

References

1. BioMed (2022), "BMC, research in progress", available at: https://www.biomedcentral.com/, (2 Jun 2022)
2. Bowers, B. (2022), "Ben Bowers", available at: https://www.gearpatrol.com/author/820004/bbowers, / (2 Jun 2022)
3. Burić, M., Pobar, M., Ivašić-Kos, M. (2018), Object detection in sports videos, In 2018 41st MIPRO Conference, IEEE, pp. 1034-1039
4. Consido Official web site, available at: https://consido.hr/
5. Datta R, Joshi D, Li J, Wang JZ. (2008), "Image Retrieval: Ideas, Influences, and Trends of the New Age", ACM Transactions on Computing Surveys, Vol.20 No.2.
6. Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., Fei-Fei, L. (2009), ImageNet: A large-scale hierarchical image database, At 2009 IEEE conference on computer vision and pattern recognition, pp. 248-255.
7. European Virtual Museum available at: https://joyofmuseums.com/museums/europe/. (2 Jun 2022)
8. Google Art Project available at: https://artsandculture.google.com/, (2 June 2022)
9. Hare, J.S. (2006), Saliency for Image Description and Retrieval, Ph.D. dissertation, Faculty of Eng. Science and Math. School of Electronics and Computer Science, University of Southampton
10. Hrga, I., Ivašić-Kos, M. (2019), Deep image captioning: An overview, In 2019 42nd International Convention MIPRO, IEEE, pp. 995-1000
11. Ivasic-Kos, M., Ipsic, I., Ribaric, S. (2015), A knowledge-based multi-layered image annotation system, Expert Systems with Applications, Vol.42 No.24, pp. 9539-9553.
12. Ivašić-Kos, M., Pavlić, M., Pobar, M. (2009), Analyzing the semantic level of outdoor image annotation. Proceedings of 32nd MIPRO 2009, Opatija
13. Ivasic-Kos, M., Pobar, M., Ipsic, I. (2014), Multi-layered Image Representation for Image Interpretation, In Proc. of the 13 Workshop on Vision and Language, pp. 115-117.
14. Kristo, M., Ivasic-Kos, M., Pobar, M. (2020), Thermal object detection in difficult weather conditions using YOLO. IEEE Access, Vol.8, pp. 125459-125476.
15. Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Lawrence Zitnick, C. (2014), Microsoft COCO: Common objects in context, in European conference on computer vision, Springer, Cham, pp. 740-755.
16. NIST a public fingerprint database, available at: https://www.nist.gov/programs-projects/nail-nail-n2n-fingerprint-capture-challenge, (2 Jun 2022)
17. PHIL (2022), "PHIL medical digital image database", available at: https://phil.cdc.gov/default.aspx, / (2 Jun 2022)
18. Photos-statistics, available at: https://phototutorial.com/photos-statistics/, (2 June 2022)
19. Pobar, M., Ivašić-Kos, M. (2015), Multimodal Image Retrieval Based on Keywords and Low-Level Image Features. In International KEYSTONE Conference on Semantic Keyword-based Search on Structured Data Sources, Springer, Cham, pp. 133-140
20. Pynck Official web site, available at: https://pynck.com
21. RadioPedia Official web site, available at: https://radiopaedia.org/
22. Rui, Y., Huang T., Chang, S. (1999), Image retrieval: Current techniques, promising directions and open issues, Journal of Visual Communication and Image Representation, Vol.10, pp.39–62.

23. Sambolek, S., Ivasic-Kos, M. (2021), Automatic person detection in search and rescue operations using deep CNN detectors. IEEE Access, Vol.9, pp. 37905-37922.
24. Siggelkow, S. (2002), Feature Histograms for Content-Based Image Retrieval, doctoral dissertation on Albert-Ludwig-Univerzitetu, Freiburg u Breisgau
25. Skimlinks, available at: https://skimlinks.com/blog/importance-of-imagery, (2 June 2022)
26. Stefanini, M., Cornia, M., Baraldi, L., Cascianelli, S., Fiameni, G., Cucchiara, R. (2021), "From show to tell: A survey on image captioning", available at https://arxiv.org/pdf/2107.06912.pdf (10 Jan 2022)
27. Van Horn, G., Aodha, O. M., Song, Y., Cui, Y., Sun, C., Shepard, A., Adam, H., Perona, P., Belongie, S. (2018), The iNaturalist Species Classification and Detection Dataset. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition.
28. Virtual tour, available at: https://virtualtours.city/, (2 June 2022)
29. Wang X, et al. (2017) ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases. IEEE CVPR 2017
30. Wang, C., Shen Y., Ji, L. (2022), "Geometry Attention Transformer with position-aware LSTMs for image captioning," Expert Systems with Applications, pp. 117174.
31. Wang, J.Z., Li, J., Wiederhold, G. (2001), "Simplicity:Semantics-sensitive integrated matching for picture libraries", Pattern Analysis and Machine Intelligence, Vol.23 No.9, pp. 947 –963.

About the author

Marina Ivasic-Kos is an Associate Professor, Dean of the Faculty of Informatics and Digital Technologies, and Head of the Laboratory for Computer Vision, Virtual and Augmented Reality at the Centre for Artificial Intelligence, University of Rijeka. She earned her Ph.D. in Computer Science at the Faculty of Electrical Engineering and Computing in Zagreb in 2012. She has been involved in numerous business and research projects in information and computer science and ICT COST, Erasmus+, and EU HKO projects. She is the leader of a national research project dealing with automatic recognition of actions in sports and a researcher at a project dealing with crowd analysis in surveillance. She also runs two projects funded by the University of Rijeka that deal with the automatic recognition of actions in sports. She received a project funded by the Science Foundation Agency for the career development of young PhDs. Her research interests focus on artificial intelligence, computer vision, soft computing, and knowledge representation. She presented her research at several international scientific conferences and in journals. She is a technical committee member and reviewer for numerous scientific conferences and reviewer for high-cited journals such as Elsevier Pattern Recognition Journal (PR), Expert Systems with Applications (ESWA) and Computers and Electronics in Agriculture, IEEE Transactions on Fuzzy Systems, Transactions on Cybernetics, Signal Processing, Access, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, MDPI Sensors and Entropy, ACM Transactions on Multimedia Computing Communications and Applications, Journal of Artificial Intelligence Research (JAIR). She is a topic board member of the MDPI Journal of Imaging. She can be contacted at email: **marina@uniri.hr.**