Enhancing photography management through automatically extracted metadata

Pedro Carvalho[1,2], Diogo Freitas[2] Tiago Machado[2], and Paula Viana[1,2]

[1] INESC TEC - Institute for Systems and Computer Engineering, Campus da FEUP, Rua Dr. Roberto Frias, 4200-465 Porto, Portugal
pedro.carvalho@inesctec.pt
[2] Polytechnic of Porto, School of Engineering, Rua Dr. António Bernardino de Almeida, 431, 4249-015 Porto, Portugal

Abstract. The tremendous increase in photographs that are captured each day by common users has been favoured by the availability of high quality devices at accessible costs, such as smartphones and digital cameras. However, the quantity of captured photos raises new challenges regarding the access and management of image repositories. This paper describes a lightweight distributed framework intended to help overcome these problems. It uses image metadata in EXIF format, already widely added to images by digital acquisition devices, and automatic facial recognition to provide management and search functionalities. Moreover, a visualization functionality using a graph-based strategy was integrated, enabling an enhanced and more interactive navigation through search results and the corresponding relations.

Keywords: Image Metadata, Exif, Face Recognition, Graph-based Visualization

1 Introduction

The acquisition of visual data has grown significantly in the last decade, supported by more accessible devices, with high quality cameras and large storage capabilities. Has a result, business as well as individuals, have amassed huge quantities of videos and images, making its management and effective usage much harder. The prevailing strategy has been the association of metadata to the images and its use with the goal of providing augmented functionalities.

Large online platforms like Google or Facebook are already exploring metadata such as date, geographic coordinates or user added information to provide some enhanced services. These require the upload of the images with all the related ownership and privacy issues. Such frameworks for local use are scarce. While metadata can be used for managing image repositories and also filter unwanted or irrelevant images, the growing quantities of visual data that are stored also demand new, more intuitive and interactive forms for visualizing the repositories, augmenting their use and value.

The work described in this paper explorers the automatic extraction and use of image metadata to (1) improve the organization and management of generic user accessible repositories and (2) provide a richer and interactive visualization over the repository. The metadata will be grounded on the EXIF standard, widely used by digital image capture devices, with additional face recognition information. Traditional search functionalities are augmented through a graph-based representation of the results, enabling a richer user interaction as well as the graphic view of the relations. The developed framework uses open software to improve flexibility and its modular architecture facilitates future enhancements.

The main contributions of this paper are: (1) exploration of EXIF metadata and of tools to enable its flexible usage; (2) base framework for management and navigation in digital repositories; (3) graph-based visualization and navigation in the repository. To our knowledge, no framework provides this set of functionalities.

The remaining of the this paper is organized as follows. Section 2 provides a brief description of relevant technologies and works. The proposed framework and corresponding architecture is described in Section 3, with metadata extraction more detailed in Sub-Sections 3.1 and 3.2. The visualization mechanisms is described in Section 4. Section 5 presents the main results and Section 6 the final conclusions.

2 Related Work

The usage of metadata has been growing over the last decades, providing an increasingly strong contribution to improve the management of content repositories and augment the value and lifespan of such data. This has been accompanied by several initiatives to define models and standards to make the usage of metadata across systems more flexible and transparent.

Manually inserted or automatic extracted metadata is being used by commercial platforms such as Facebook, Instagram or Google Photos. The later provides some features to improve image searches, using data present in the photos (e.g., location, time or landscapes) to group them. Facebook users create metadata when they manage friend lists, post status or add descriptions to content. By tracking these activities, Facebook analyses trending topics and promotes sponsored posts that generate revenue. Instagram users provide captions to the images and follow other users; this interaction data is used to improve advertising.

Some metadata models or standards have focused on specific scenarios or application areas, like EXIF [10] or IPTC [6] for photography, while others, such as MPEG-7 [2], have tried to be more universal, but this generalization as also impaired its adoption.

Dublin Core metadata [1, 9] is a standard that defines metadata for networked electronic information and core elements such as contributor, coverage, creator, date, description, format, identifier, language, publisher, relation, rights, source, subject or title. The IPTC model [6] was designed to capture important information to news gathering, reporting and publishing activities; the usage of this metadata, embedded in image file formats, became widespread with the use of the Adobe Photoshop. Adobe also developed the Extensible Metadata Platform (XMP) standard [7] for the creation, processing and exchange of metadata in a variety of applications. XMP uses the Resource Description Framework (RDF) technology for data modelling and defines how the data model is serialised and incorporated into an image file. MPEG-7 [2], is one of the standards from the MPEG family; it is not aimed at any specific application and defines how to describe different aspects related to content, including format, content or devices. However, this generalization, is subsequent complexity, as also impaired its adoption. The Exchangeable Image File Format (EXIF) [10] is a image file storage standard for digital cameras. It was developed to store images created by digital cameras, as well as metadata about the images and is currently widely used in many other devices, including smartphones.

More recently, deep learning-based models are being extensively explored for object detection and recognition, enabling the automatic extraction and association of metadata to multimedia content, with face recognition playing an important role [4].Two main approaches have been used for facial recognition Convolutional Neural Networks (CNN) [5]; Siamese Networks [3]. The former identifies region proposals which are then classified based on a set of know faces used in the training stage, while the later uses two symmetric neural networks that extract embedding of the faces to be recognized and from a database of faces, to compute the distance between the embeddings to determine their similarity.

3 Proposed Framework

The architecture of the proposed framework is depicted in Fig 1. The web graphic interface is intended for user interaction, allowing information and images to be upload, searched and visualized. The database is used for the organised storage of metadata information, image management and supports search functionalities. The image files themselves are stored in the repository. The central Server manages the overall process and performs metadata extraction.

The automatic extraction of metadata plays a key role in the proposed framework, with two main types being used: EXIF metadata; facial recognition. When an image is uploaded it is stored in the repository and EXIF metadata, if present, is extracted and parsed to be stored in the database; in parallel, face detection modules are used to identify the corresponding regions in the image an compute the face embedding, avoiding the storage of the faces as images.

For implementation, SQLlite was used for the database. The user interface was implemented using web technologies, making its usage in different devices more flexible. The back-end was implemented using Python with the communication between GUI and server using Web Services.

Subsection 3.1 justifies the usage of EXIF metadata and the process that was adopted for its extraction, while subsection 3.2 describes the strategy for facial detection and comparison.

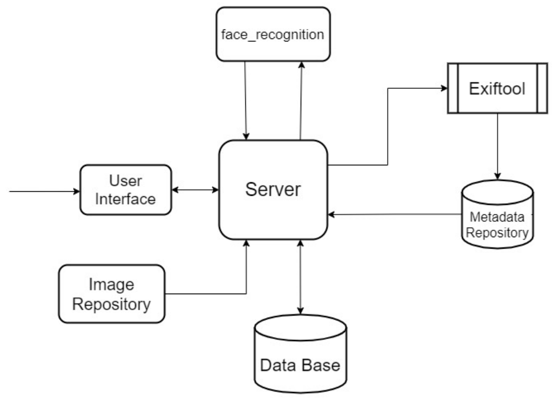


Fig. 1: Schematic of the system architecture.

3.1 EXIF Metadata

At the current stage of development, the framework relies on EXIF image metadata as a basis. Image metadata refers to additional information about images, which is also stored in image files. EXIF has been chosen for this work due to its widespread automatic use in acquisition devices including digital cameras and smartphones, which are intensively used daily by people worldwide. Hence, this framework is suitable for most users who use their smartphones or personal cameras for their average day photographs, as well as professional photographers. From the many fields defined in EXIF, at this phase only a subset was chosen to be incorporated in the framework, which includes: date; time; geographic coordinates; image information; user comments. The inclusion of image information is intended to provide knowledge about characteristics of the image such as formats and resolution. The other mentioned fields were selected to enable the identification of relations between images rather than just the use of keywords. Other fields may be added in later versions.

After the image is uploaded, it is analysed and the image metadata is extracted through the use of the Exiftool software (https://exiftool.org) that was integrated in the framework. Due to its characteristics, it is executed through a system call using Python's "OS" module. One of the functionalities provided by Exiftool enables saving all the data extracted from an image in a text file with the same name as the image. For that, it requires the modification of the executable name from "exiftool(-k).exe" (original name) to "exiftool (-a -u -g1 -w txt).exe".

Exiftool extracts all the EXIF metadata present, requiring an additional parsing of the generated text file to filter the desired tags and the subsequent storage of the the information in the database. This search is done using the tag names, as defined by the EXIF standard. To optimise the process, if some of the information is already present in the database, such as image format or resolution, only a reference is added in the corresponding image identification entry. Otherwise, the new information is added to the database before adding the reference.

3.2 Face Detection and Recognition

To enhance metadata extraction, a human face detection and identification module was integrated. It uses a detection method based on HOG (Histogram of Oriented Gradients) for face detection; HOG features are extracted and used as input to a SVM (Support Vector Machine) classifier, previously trained for face detection, that identifies the image regions likely to contain faces. These regions are then used to extract encodings from the faces based on the method developed by Kazemi et al. [8]; using this method 68 reference points are identified. This is used as input to a deep neural network, whose main function is to generate an embedding with 128 measurements, based on the reference points. The comparison of embeddings is then used to recognise faces. When an image is uploaded, the embedding for each detected face is computed and only these are stored; hence, no image data from the faces is stored and the performance of the search process is also favoured. Fig. 2 illustrates the mapping of a face into embeddings.

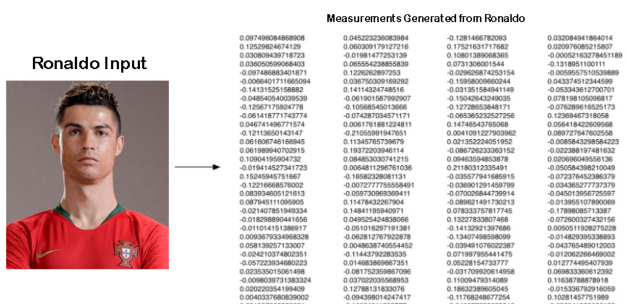


Fig. 2: Illustration of the generation of the 128 measurements for face embedding.

4 Image Search and Visualization

Traditional search and visualization mechanisms, such as search text tool and mosaic presentation, have been enhanced through the implementation of a new feature that enables a more interactive usage. This new graphic element is based on graphs and was added to improve the usability and navigation trough search results and the content of the repository, overcoming limitations caused by a large number of images to display.

Network diagrams, also known as graphs, are diagrams that show interconnections between a set of entities, symbolising that they are related in some way to each other. Each entity is represented by a node (or vertex). The connections between nodes are represented by links (or edges). Using graphs enables the presentation of image information associating it to the nodes, as well as the representation of relations between the images through the edges connecting nodes. This representation favours highlighting the most relevant factors.

The graph-based view was integrated in the web interface through the use of the D3.js module (https://d3js.org/). In this project, the entities (or nodes) will be images, with selected images playing a central role, and the remaining nodes are related images. Then, depending on the strength of the relation between nodes, the connecting edges will be thinner or thicker, being weaker or stronger, respectively.

The use of the D3.js library, more specifically graphs, in this project consists of presenting the existing relationships between the various images submitted. In order for this to be possible, it was necessary to access the necessary information in an organized manner. It was necessary to perform the necessary implementations to map the metadata about the images that is stored in the database into the appropriate input for the D3 tool. Specifically, a JSON file is used that includes the following key topics. Listing 1.1 depicts an excerpt of the JSON file, illustrating how information is mapped.

Listing 1.1: Example of JSON file structure to build D3 graph.

```
{
    "links":[
        {
            "source": SourceNodeID,
            "target": TargetNodeID,
            "value": value for connection's strength
        }
    ],
    "nodes":[
        {
            "group": group in which the node will be presented
            "id": node identifier
        }
    ]
}
```

5 Results

A set of tests were conducted to assess the performance of different aspects of the system, the first ones focusing on facial recognition. Even though, the performance of the face detection and recognition module as been tested in previous works, its integration in the framework, was analysed: (1) with images without faces to assess false positives; (2) image with one face to assess the ability to find other images with the same person; (3) photos with several faces to assess the search for images with all or some of the detected individuals.

In the first test, when submitting or selecting images without faces, the image detection module correctly does not detect anything; hence, no images are retrieved based on this information, requiring additional information to establish relations with images in the repository. Fig 3 depicts an example of the graphic representation of the relations between images; the metadata used consisted of: date and time of creation; model of the device; user accounts.

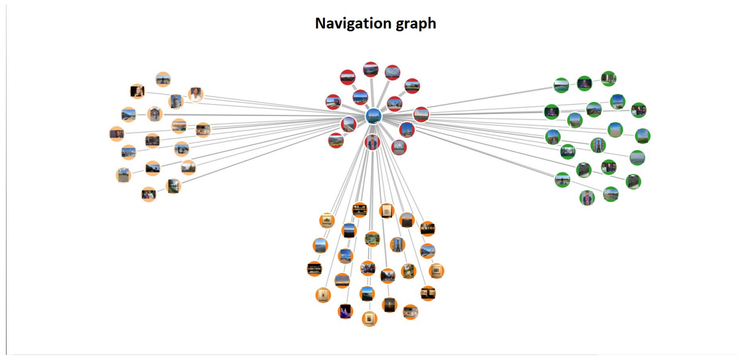


Fig. 3: Graphic representation of selected images (in the central group) and searched images based on selected EXIF metadata..

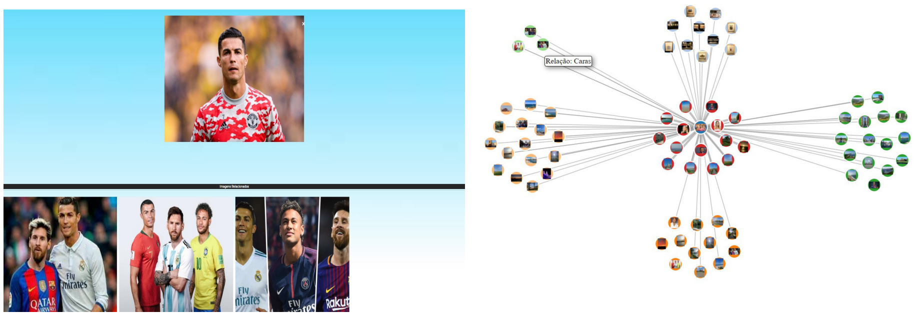
Fig. 4 illustrates results of the second test where an image with a single face is used to search the repository; here, in addition to the previous metadata fields, face recognition is also used. Fig. 4a shows the selected image on top and matched images bellow. The corresponding graph view is depicted in Fig. 4b, where it is visible the new group due to face recognition on the top left corner.

For the final test, an image with multiple faces is used as input for the search (Fig. 5). An example of such image is depicted in Fig 5a along with possible matches. The corresponding graph is depicted in Fig. 5b.

6 Conclusion

This paper presents a small and lightweight framework that explores image metadata to assist the management and navigation though image repositories. The framework uses EXIF metadata present in images, widely used in smartphones and digital cameras, augmented by facial recognition. A graph-based visualization mechanism was integrated to provide a graphic and interactive visualization and navigation through the repository.

The modular and distributed architecture, with the communication built upon web services, enables a flexible growth as well as the execution of the framework in a single computer or distributed in two or more machines. Future improvements may include the integration of additional EXIF metadata fields, as well as additional image processing algorithms for information extraction such



(a) Screenshot with selected image with only one face (on top) and searched images bellow.

(b) Graphic representation of search results including face recognition; images search based on face recognition on the top left group.

Fig. 4: Relationships of images in the repository with an image with a single face.

(a) Screenshot with selected image with several faces (on top) and searched images bellow.

(b) Graphic representation of search results including face recognition; images search based on face recognition on the top left group.

Fig. 5: Relationships of images in the repository with an image with multiple faces.

as object recognition. Also, functionalities of D3, or other graph-based tools, should be further explore to enhance even more the visualization capabilities.

Acknowledgement

The work presented was financed by National Funds through the Portuguese funding agency, FCT—Fundação para a Ciência e a Tecnologia, within project UIDB/50014/2020

References

1. DCMI Usage Board. Dublin core metadata element set, version 1.1: Reference description, 2012. https://www.dublincore.org/specifications/dublin-core/dces/.
2. Shih-Fu Chang, T. Sikora, and A. Purl. Overview of the mpeg-7 standard. *IEEE Transactions on Circuits and Systems for Video Technology*, 11(6):688–695, 2001.
3. D. Chicco. Siamese neural networks: An overviews. *Artificial Neural Networks. Methods in Molecular Biology*, 2190, 2021.
4. IBM Cloud Education. Deep learning, 2020. IBM at https://www.ibm.com/cloud/learn/deep-learning.
5. Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. http://www.deeplearningbook.org.
6. IPTC Photo Metadata Working Group. Iptc photo metadata standard 2019.1, 2020. http://www.iptc.org/std/photometadata/specification/IPTC-PhotoMetadata.
7. ISO. Graphic technology — extensible metadata platform (xmp) — part 1: Data model, serialization and core properties, 2019. https://www.iso.org/obp/ui/#iso:std:75163:en.
8. Vahid Kazemi and Josephine Sullivan. One millisecond face alignment with an ensemble of regression trees. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1867–1874, 2014.
9. NISO. Understanding metadata, 2004. https://www.lter.uaf.edu/metadata_files/UnderstandingMetadata.pdf.
10. Technical Standardization Committee on AV, IT Storage Systems, and Equipment. Exchangeable image file format for digital still cameras: Exif version 2.2, 2002. PDF at https://www.exif.org/Exif2-2.PDF.