

IMGpedia: Enriching the Web of Data with Image Content Analysis

Sebastián Ferrada, Benjamin Bustos, and Aidan Hogan

Center for Semantic Web Research
Department of Computer Science
University of Chile
{sferrada, bebustos, ahogan}@dcc.uchile.cl

Abstract. Linked Data rarely takes into account multimedia content, which forms a central part of the Web. To explore the combination of Linked Data and multimedia, we are developing IMGpedia: we compute content-based descriptors for images used in WIKIPEDIA articles and subsequently propose to link these descriptions with legacy encyclopaedic knowledge-bases such as DBPEDIA and WIKIDATA. On top of this extended knowledge-base, our goal is to consider a unified query system that accesses both the encyclopaedic data and the image data. We could also consider enhancing the encyclopaedic knowledge based on rules applied to co-occurring entities in images, or content-based analysis, for example. Abstracting away from IMGpedia, we explore generic methods by which the content of images on the Web can be described in a standard way and can be considered as first-class citizens on the Web of Data, allowing, for example, for combining structured queries with image similarity search. This short paper thus describes ongoing work on IMGpedia, with focus on image descriptors.

1 Introduction

WIKIPEDIA centres around the curation of human-readable encyclopaedia articles, where at the time of writing it contains about 38 million articles in almost 200 languages. In terms of structured content, the DBPEDIA project [3] systematically extracts meta-data from WIKIPEDIA articles and presents it as RDF. Another initiative, called WIKIDATA [8] – organised by the Wikimedia Foundation – allows users to directly create and curate machine-readable encyclopaedic entries. In terms of multimedia content, WIKIMEDIA COMMONS¹ is a collection of 30 million freely-usable media files (image, audio, video) that are used within WIKIPEDIA articles. Recently, DBPEDIA COMMONS [7] was released: a knowledge-base that takes the meta-data from each multimedia file’s description page – such as the author, size and licensing of the media – and publishes it as RDF. However, none of the existing knowledge-bases in this space describe aspects of the multimedia *content* itself.

Because of this we are developing IMGpedia [1], a new knowledge-base that uses meta-data from DBPEDIA COMMONS and enriches it with *visual descriptors* of all images from the WIKIMEDIA COMMONS dataset. The goal is to allow users to perform “visuo-semantic queries” using IMGpedia, e.g., *retrieve a painting for every European painter from the 17th century* or *given a photo of yourself, obtain the top-k most similar portraits painted by an American artist*. We could also infer new knowledge about DBPEDIA entities, given the relations between the images;

¹ <https://commons.wikimedia.org>

e.g, if two entities of type `dbo:Person` share an image (or they are very similar), we could infer, with some given likelihood, that they have met. Our general goal is to investigate methods by which descriptors of the content of images – not just their meta-data – can be integrated with the Web of Data in a standard way.

In previous work [1], we proposed the general goals of IMGPEdia. This paper is a brief progress update: currently we are downloading the images from WIKIMEDIA, computing the visual descriptors for each image, and creating the image entities that will be linked with existing encyclopaedic knowledge-bases. We thus propose a set of existing multimedia algorithms for extracting descriptors from images, describe their applications, and provide reference implementations in various programming languages such that they be used as a standard way to (begin to) describe images on the Web of Data. We provide some initial results on computing descriptors for the images of WIKIMEDIA COMMONS, we describe some of the technical engineering challenges faced while realising IMGPEdia, and we outline future plans.

2 Reference Implementations for Visual Descriptors

Aside from simply providing the URL of an image, we wish to compute visual descriptors for each image. These descriptors are vectors produced after performing different operations over the matrix of pixels in order to obtain certain features, such as color distribution, shapes and textures. Later, those vectors are stored as a part of a metric space in which similarity search can be performed. We are considering four different image descriptors which we have already implemented in Java, Python and C++, and we have proven to give equivalent results among different languages over a dataset of 2800 Flickr images. The first three descriptors require a preprocessing step, which is to convert the image to greyscale using intensity $Y = 0.299 \cdot R + 0.587 \cdot G + 0.114 \cdot B$, where R , G and B are the values of each color channel: red, green and blue, respectively. The four descriptors we compute are:

Gray Histogram Descriptor: images are partitioned in a fixed amount of blocks.

Per each block a histogram of gray intensity is computed; typically intensity takes 8 bit values. Finally, the descriptor is the concatenation of all histograms.

Oriented Gradients Descriptor: image gradient is calculated via convolution with Sobel kernels [6]. A threshold is then applied to the gradient, and for those pixels that exceed it, the orientation of the gradient is calculated. Finally, a histogram of the orientations is computed and is used as the descriptor.

Edge Histogram Descriptor: for each 2×2 pixel block, the dominant edge orientation is computed (horizontal, vertical, both diagonals or none), where the descriptor is a histogram of these orientations [4].

DeCAF7: a Caffe neural network pre-trained with the Imagenet dataset [2] is used. To obtain the vector, each image is resized and ten overlapping patches are extracted; each patch is given as input to the neural network and the last self-convolutional layer of the model is extracted as a descriptor, so the final vector is the average of the layers for all the patches [5].

The implementations and their dependency/compilation documentation can be found at <https://github.com/scferrada/imgpedia> under GNU GPL license.

3 Performance Evaluation of Descriptors

In order to build the knowledge-base, preliminary steps must be done. First of all, it is necessary to obtain a local copy of the WIKIMEDIA COMMONS image dataset

Table 1. Descriptor calculation benchmark

Descriptor	Average time per image [ms]	
	Single Thread	Multithread
GHD	119.4	10.4
HOG	477.7	25.7
EHD	729.1	42.1
DeCAF7	3634.8	3809.7

so we can keep IMGPEdia updated in the future. To do so, WIKIMEDIA provides an *rsync*-protocol server for downloading the ~ 21 TB of images of the dataset. Currently we have downloaded about ~ 19 TB at an average speed of ~ 550 GB per day, corresponding to ~ 14 million images downloaded from a total of ~ 16 million. Each image is stored locally, along with its four descriptors.

We benchmarked the computation of descriptors in order to know the time such extraction will take. Experiments were performed on a machine with Debian 4.1.1, a 2.2 GHz (24 core) Intel(r) Xeon processor, and 120 GB of RAM. We computed the descriptors for a sub-folder of the WIKIMEDIA COMMONS dataset, containing 57,377 images. The process is comprised of three steps: read the image and load it into memory, extract the descriptor itself and save the vector on disk. Two experiments were performed, one using a single execution thread and the other using 28 threads, where each thread handles an image at a time. The neural network for DeCAF7 was used in CPU mode (rather than GPU mode).

Results are shown in Table 1, where we see that DeCAF7 is the most expensive descriptor to run. While we can see an improvement of one order of magnitude in calculation time for the first three descriptors, DeCAF7 does not benefit: the Caffe implementation for neural networks already uses multithreading for its computation on a single image, so assigning different images to different threads gives a slight *overhead* since multiple cores are already being exploited. On the present hardware, we estimate it would take ~ 1.8 years to run DeCAF7 over all 16 million images; hence, we will have to select a subset of images on which to run this descriptor.

4 Next Steps

Linking IMGPEdia with DBPEdia, WIKIDATA, etc. To complete our knowledge-base and be able to infer new relations, we must combine these visual descriptors with the semantic entities of DBPEdia and WIKIDATA. For this, we must extract the links between the articles and the images that they use since they are not present in DBPEdia COMMONS and only present in DBPEdia for some images. This is an ongoing effort in which we are researching the possible options to perform this task: we hope to exploit the WIKIMEDIA dumps, either to derive this information directly, or to parse it from the articles themselves. Subsequent to this, we wish to investigate using descriptor-based user-defined-functions in SPARQL to enable visuo-semantic queries that combine, e.g., image similarity with knowledge-base queries.

Efficient algorithms for finding relations between similar images To facilitate querying IMGPEdia, we propose to compute static relationships between images; e.g., if two images have a descriptor distance that tends to 0, then we can state that those

images are `nearCopy` of each other. Other distance-based relations can be computed based on other fixed thresholds, such as `similar`. The brute-force approach for finding all pairs of images within a given distance takes quadratic comparisons. Thus, the challenge is to do this efficiently, where we propose to explore: building an index structure for similarity search over the dataset; using approximate similarity search algorithms; using self-similarity join techniques; and so forth.

Labelling images for multimedia applications By linking `IMGpedia` to existing knowledge-bases, we hope to label images with categories, types, entities, etc. Thus `IMGpedia` could serve as a useful resource for the multimedia community; e.g., since we are using Caffe framework for neural networks to compute DeCAF7 [2], we could train a network using the `DBpedia COMMONS` categories extracted for each image as labels, allowing us to automatically classify new images in the dataset or benchmark performance against other classifiers. We could also label images with the specific entities they contain using `DBpedia/Wikidata` based on the article(s) in which it appears: we could need to tread careful since, e.g., an image used on an article of a person may not contain that person, but we could consider some heuristics such as the location of the image and some further content analysis.

5 Conclusions

In this short paper, we have given updates of our ongoing work on `IMGpedia`. We benchmarked four image descriptors for which reference implementations are made public. We discussed a number of open engineering challenges, as well as some open research questions we would like to study once the knowledge-base is completed.

Acknowledgments This work was supported by the Millennium Nucleus Center for Semantic Web Research, Grant N^o NC120004, and Fondecyt, Grant N^o 11140900. We thank Camila Faundez for her assistance.

References

1. Bustos, B., Hogan, A.: `IMGpedia`: a proposal to enrich `DBpedia` with image meta-data. In: AMW. vol. 1378, pp. 35–39. CEUR (2015)
2. Donahue, J., Jia, Y., Vinyals, O., Hoffman, J., Zhang, N., Tzeng, E., Darrell, T.: `Decaf`: A deep convolutional activation feature for generic visual recognition. arXiv preprint arXiv:1310.1531 (2013)
3. Lehmann, J., Isele, R., Jakob, M., Jentzsch, A., Kontokostas, D., Mendes, P.N., Hellmann, S., Morsey, M., van Kleef, P., Auer, S., Bizer, C.: `DBpedia - A Large-scale, Multilingual Knowledge Base Extracted from Wikipedia`. *Semantic Web Journal* (2014)
4. Manjunath, B.S., Ohm, J.R., Vasudevan, V.V., Yamada, A.: Color and texture descriptors. *IEEE Trans. on Circuits and Systems for Video Tech.* 11(6), 703–715 (2001)
5. Novak, D., Batko, M., Zezula, P.: Large-scale image retrieval using neural net descriptors. In: *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*. pp. 1039–1040. ACM (2015)
6. Sobel, I., Feldman, G.: A 3×3 isotropic gradient operator for image processing. *Stanford Artificial Project* (1968)
7. Vaidya, G., Kontokostas, D., Knuth, M., Lehmann, J., Hellmann, S.: `Dbpedia commons: Structured multimedia metadata from the wikimedia commons`. In: *The Semantic Web-ISWC 2015*, pp. 281–289. Springer (2015)
8. Vrandečić, D., Krötzsch, M.: `Wikidata: A free collaborative knowledgebase`. *Comm. ACM* 57, 78–85 (2014)