Non-textual Data Extraction Assigment

Content Based Image Retrieval for cars images

Julio Martínez Bastida, Carlos Sánchez Velázquez
and Rafael Sojo García

08/03/2023

Contents

1 Introduction and Problem Description

With the advance of technologies, the possibility of analysing big amounts of digital images increases, and so different challenges associated to it. In order to face these challenges, different techniques such as Information extraction appear. Information extraction in images refers to the process of obtaining the relevant information from an image and representing it in a way that helps to make comparisons and image retrievals. To accomplish this is common to see the use of image descriptors, which are mathematical representations that capture the key visual features of an image.

One possible practical application of images information extraction is the task of identifying similar cars to a given image. This task involves analysing visual features of the car, for example its shape, colour, and texture, and afterwards comparing them with a database of car images to find other images that are visually similar and try to identify the model of the car. This task has several applications in various real domains.

For example, in the main commercial car industry can benefit from this technology, they can use it to detect or find cars similar to the ones their clients could be interested in buying and maybe recommend them, increasing the possibility of accomplishing a sale.

Another industry that could beneficiary from this type of technology could be the insurance industry, where it can be used to assess damage to vehicles involved in accidents. These companies can identify in a fast way the model of a damaged car and compare it with a database of images to determine the extent of the damage and the estimated cost of repairs.

CBIR stands for Content-Based Image Retrieval. It is a technique used to search and retrieve images from a large collection of images based on their visual content. In this project, our goal is to develop a CBIR system for car images, which identifies similar cars to a given image.

The choice of using a CBIR for this task is evident. CBIR systems have the main characteristics needed for a task of this type, they use visual features extracted from the images, such as colour or texture which means that the system can effectively compare the visual characteristics of cars and find the most similar cars to the query image. Furthermore, CBIR systems can handle large databases of images, which is useful for this type of tasks, since there are many different makes and models of cars on the market that should be stored in a dataset.

To put in practice this system we selected a dataset composed by more than 16200 images of different cars by different angles extracted from Kaggle [1].

2 State of the art

Content-based image retrieval aims to respond to a query as precise as possible with images and was introduced by Chang [2] and that since the rise of the use of neural networks has generated great expectations, as Wan et al.[3] mention.

However, in this paper we will take a humbler approach, using ORB descriptors for the image analysis, and the Hamming distance to compare between descriptors of different images. After that we will use the histograms to filter between the best results of the previous analysis using the χ distance to compare the different channels of the different colour spaces we are considering.

2.1 Keypoints descriptors

Keypoint descriptors are a crucial component of content-based image retrieval (CBIR) systems. They enable the extraction of distinctive features from images, which can be compared to identify similarities between images. Over the years, several keypoint descriptor methods have been proposed and developed, each with their own strengths and weaknesses.

One of the earliest and most widely used keypoint descriptor methods is the Scale-Invariant Feature Transform (SIFT) algorithm, proposed by Lowe in 1999 [4], which is known for its robustness to changes in illumination, scale, and orientation. However, SIFT is computationally expensive, which limits its scalability.

Due to that last fact, a method called Speeded Up Robust Features (SURF) was proposed by Bay et al in 2006 [5]. SURF is an improvement over SIFT, using a faster algorithm for detecting keypoint locations and a more efficient descriptor representation being at the same time also robust to scale, rotation, and illumination changes, making it suitable for real-world applications.

Other algorithms were proposed with the objective of being faster and more efficient than previous ones but with the drawback of being less robust to scale and rotation changes, examples of these methods are Features from Accelerated Segment Test (FAST) algorithm, proposed by Rosten and Drummond in 2006 [6], the Binary Robust Independent Elementary Features (BRIEF) algorithm, proposed by Calonder et al in 2010 [7], or Binary Robust Invariant Scalable Keypoints (BRISK) algorithm, proposed by Leutenegger et al. in 2010 [8].

In this case, we decided to use the Oriented FAST and Rotated BRIEF (ORB) algorithm, proposed by Rublee et al [9]. in 2011, which is another popular keypoint descriptor method. ORB combines the FAST corner detection algorithm taking advantage of its computational properties, with the BRIEF descriptor representation using a rotation-invariant version of BRIEF to achieve orientation invariance. ORB is designed to be fast and efficient, making it suitable for large scale CBIR systems.

To generate the ORB descriptors, first the FAST algorithm is used to detect corner points in an image, which are potential keypoints. Then, the BRIEF descriptor is computed at each FAST keypoint to generate a binary feature vector that encodes the local image information around the keypoint.

In CBIR systems, various distance measures can be used to compare keypoints and match them to measure similarities among images. Distances such as Euclidean distance and Manhattan distance are used with continuous-valued feature vectors. In this case, the Hamming distance was chose, which is used to compare binary feature vectors, such as those generated by ORB. It computes the number of bits that are different between two binary vectors, which provides a measure of the dissimilarity between them. This distance can identify similar keypoints in an efficient way

2.2 Color histograms descriptors

For the histogram analysis part, we first opted to use a K-means based model based on [10]. However, after further study of the use case, we needed to obtain the most similar images, so we discarded the idea. In addition, OpenCV was able to discretise the channels of the colour spaces, so we only required a distance to compare between the query image and the catalogue images. It is also mentioned in numerous articles, such as [11] and [12], that the χ^2 distance generates good results when few bins are used, so we opted to use 8.

On the other hand, we have chosen 3 different color spaces from which 3 different histograms were obtained at each one of them. Thus, covering the most important descriptive aspect of an image. The selected channels were RGB, HSV and CIE Lab, and there are several reasons behind this selection. First, the RGB channel is the most widely use for representation and covers the 3 main primary colors, however, just these 3 channels are not enough considering all the details that a human eye can differentiate. For that reason, with the HSV color space we make sure that all the chromatic circle is covered within the Hue channel, while the saturation is also taken into account. [13]. Then, the lightness and brightness of the image can be measured with the L and V channels from CIE Lab and HSV respectively, where despite the fact that both channels are measuring the luminosity of the image, this is done in different ways. Finally, the a and b channels are covering the difference in color from red to green, and yellow to blue respectively, with a bit more detail than the Hue channel. [14].

3 Implementation

Our approach is very simple although it accomplish its goal pretty well. One of the mandatory requirements of this project was to use an histogram similarity measure for the implementation, however, this approach does not take into account the differences in shape between the cars of two images. Due to this fact, as it was mentioned in the introduction, we decided to use the ORB descriptor as a first step in the process of ranking the cars, and only then the histogram similarity is computed. The process of matching goes as follows:

1. Histograms and Keypoints descriptors are calculated for each image.
2. A set of matching Keypoints are obtained from the comparison between the reference image and each one from database using the Hamming norm.
3. The mean of distances between matching Keypoints is computed for each image
4. The list of mean distances at each comparison is sorted, from where the 10 sorter distances are selected.
5. A similarity score between histograms is obtained using the χ^2 metric.
6. The mean of similarity scores between histograms for all the color spaces is computed.
7. The new list of mean distances is again sorted, from which the best 5 scores are selected.

This set is then considered by the algorithm as the list of best matching images for the input selection. Here, there are some examples.

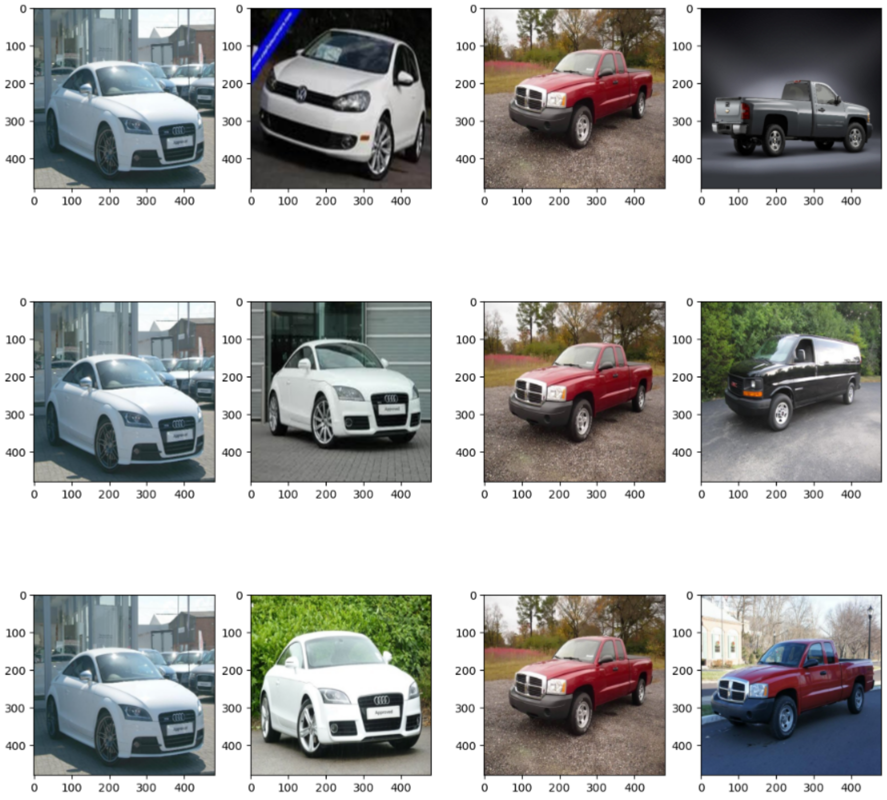

Figure 1: Results for two reference images in columns 1 and 3

The database of images that we have used for the comparison has about 8000 images and the process of calculating Histograms and Keypoints for each one of them takes about 2 minutes. For that reason, we decided to create a metadata folder in which a JSON file containing the image information is stored. Later, the process of reading takes just 1 minute, which a considerable reduction in cost terms for the computer. Finally, it is worth to mention that the matching process takes a few second for each image to give the results. Of course, all the process could be much faster with a correct parallelization, but considering the time we had it was set for future lines of work.

4 Conclusion

To conclude this project, we can affirm that the matching process between images is not an easy task since there are many aspect to consider in terms of color, shape and texture, whereas the selection of the correct distance metrics also plays an important role. However, we can say that in some situations the database of images is as important as those metrics, since during the development of the project we find out how some specific images with perhaps better quality, managed to scale positions within the ranking recurrently. In any case, we achieved relatively good results.

5 References

References

[1] J. Li, "Stanford cars dataset." Retrieved from `https://www.kaggle.com/datasets/jessicali9530/stanford-cars-dataset`, 2018.

[2] S. Chang and S. Liu, "Picture indexing and abstraction techniques for pictorial databases," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, no. 4, pp. 475–484, 1984.

[3] Y. Liu, D. Zhang, and G. Lu, "A survey of content-based image retrieval with high-level semantics," *Pattern Recognition,(January 2007).*

[4] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[5] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," *European Conference on Computer Vision*, p. 404–417, 2006.

[6] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," *European Conference on Computer Vision*, p. 430–443, 2006.

[7] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "Brief: Binary robust independent elementary features," *European Conference on Computer Vision*, p. 778–792, 2010.

[8] S. Leutenegger, M. Chli, and R. Y. Siegwart, "Brisk: Binary robust invariant scalable keypoints," in *2011 International Conference on Computer Vision*, pp. 2548–2555, 2011.

[9] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "Orb: an efficient alternative to sift or surf," *International Conference on Computer Vision*, pp. 2564–2571, 2011.

[10] C. Lin, R. Chen, and Y. Chan, "A smart content-based image retrieval system based on color and texture feature," *Image and vision Computing*, vol. 27, no. 6, pp. 658–665, 2009.

[11] O. Pele and M. Werman, "The quadratic-chi histogram distance family," in *Computer Vision–ECCV 2010: 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5-11, 2010, Proceedings, Part II 11*, pp. 749–762, Springer, 2010.

[12] M. Roederer, W. Moore, A. Treister, R. R. Hardy, and L. A. Herzenberg, "Probability binning comparison: a metric for quantitating multivariate distribution differences," *Cytometry: The Journal of the International Society for Analytical Cytology*, vol. 45, no. 1, pp. 47–55, 2001.

[13] "HSL and HSV - Wikipedia — en.wikipedia.org." `https://en.wikipedia.org/wiki/HSL_and_HSV`. [Accessed 08-Mar-2023].

[14] "Espacio de color Lab - Wikipedia, la enciclopedia libre — es.wikipedia.org." `https://es.wikipedia.org/wiki/Espacio_de_color_Lab`. [Accessed 08-Mar-2023].