Visual Recommendation Use Case for an Online Marketplace Platform - allegro.pl

Anna Wróblewska
Allegro Group
Warszawa, Poland
anna.wroblewska@allegrogroup.com

Łukasz Rączkowski
Allegro Group
University of Warsaw
Toruń, Poland
lukasz.raczkowski@allegrogroup.com

ABSTRACT

In this paper we describe a small content-based visual recommendation project built as part of the Allegro online marketplace platform. We extracted relevant data only from images, as they are inherently better at capturing visual attributes than textual offer descriptions. We used several image descriptors to extract color and texture information in order to find visually similar items. We tested our results against available textual offer tags and also asked human users to subjectively assess the precision. Finally, we deployed the solution to our platform.

Keywords

visual search; image search; content-based recommendations; e-commerce; online auctions; image processing; LIRE; OpenCV; Elasticsearch; CBIR

1. INTRODUCTION

Content based image retrieval (CBIR) has slowly but steadily been growing in importance as a means to improve search experience in e-commerce. Several prominent companies have already deployed their own image retrieval solutions. Pinterest developed an image search platform and thus they showed that content recommendation powered by visual search improves user engagement [10]. At eBay it's been proven that image-based information can be used to quantify image similarity, which can be used to discern products with different visual appearances [2]. Finally, Yahoo built a visual similarity-based interactive search system, which led to more refined product recommendations [8].

At Allegro [1] our main goal is the satisfaction of our users, so in order to further push for that objective, we've decided to try our hand at visual recommendations. The aim of the project was to develop a system that returns product offers that are visually similar to the given input image. That similarity is determined based on several features extracted from offer images. Low-level image information like that is particularly useful in product categories for which visual aspects are important and also for categories that contain products which are hard to describe with words, e.g. fashion or jewellery items.

We started our project in a very small agile team consisting of a data scientist, a software engineer, a UX designer and a product owner. From the very start our main objective was to create a data-driven product and deploy it within the Allegro platform. We were given a strict deadline to deliver business value, so we've decided to utilize readily available open-source projects.

In the following sections, we describe these topics in detail: presentation of available offer data stored in the Allegro platform (section 2), software architecture of our solution, used tools, image features and similarity metrics (section 3), variety of techniques to test the solution based on available data (section 4). In the last section we conclude the paper and discuss future work.

2. CHARACTERISTICS OF OFFER DESCRIPTIONS

Offers listed in our marketplace platform can be described by a limited set of well-defined attributes, short title and a long description that contains unstructured data with a lot of additional information, e.g. seller addresses or tailored recommendations. Each offer is categorized and each should include at least one photo that shows the offered product.

Usually text parameters of offers in our service are high-level abstractions and as such are not very precise, e.g. a dotted pattern may describe a large spectrum of dot sizes and configurations. Furthermore, our traditional textual color tags have only around 10 very general values, e.g. shades of yellow or shades of brown. In spite of the fact that color and pattern attributes are obligatory, our sellers often specify them ambiguously as "other color", "other pattern" or "multicolor". Thus recommendations based only on text attributes are sometimes quite vague and not precise. A cursory data analysis in a single fashion category (Figure 1) shows that a large fraction of fashion items listed online lacks precise color information.

Having such a huge disparity in attribute quality drove us to hypothesize that by extracting image features we can get a better understanding of offers in our service.

The quality of images attached to product offers varies greatly. In some categories they're quite bad, and in others they're very professional. To overcome this quality hurdle in the beginning of our work, we've decided to test our algorithms only on images from a special part of Allegro called Brand Zone. These images usually have a bright background and the main object/product is in the middle. Thanks to this we were able to easily crop the essential part of every image.

We came to this conclusion last year, when we tested the offer image quality in a single fashion product category using our dedicated tool – Nigel - that marks image quality on the scale from 0 (bad quality) to 1 (the best quality). In Figure 2 we show a

SIGIR'16, July 17–21, 2016, Pisa, Italy.
© 2016 ACM. ISBN 978-1-4503-4069-4/16/07…$15.00.
DOI: http://dx.doi.org/10.1145/2911451.2926722

density distribution plot that showcases how Nigel scores images in the Brand Zone versus Allegro as a whole.

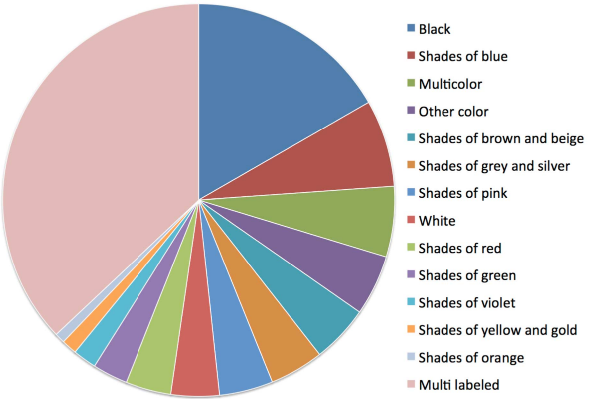

Figure 1. Color attributes distribution in the dresses category. Almost 50% of offers have multi-labeled assignments or values like multicolor / other color.

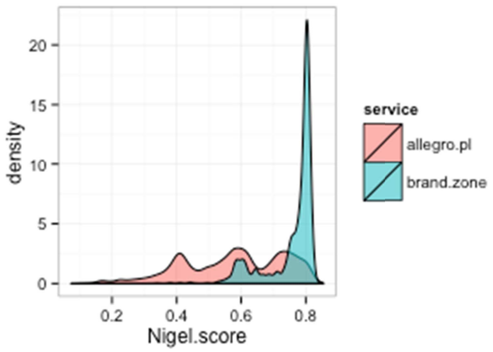

Figure 2. Density distribution of image quality assessment with the Nigel tool.

3. TOOLS AND METHODS

3.1 Solution Architecture

We developed our product in two stages. First, we created the necessary backend for the image retrieval system and also a test web application, which we affectionately named Imagator (Figure 3). In the development of the backend we utilized the following open-source projects: Elasticsearch (ES) [6], LIRE (Lucene Image Retrieval) [11], OpenCV [4] and elasticsearch-image [5]. Elasiticsearch serves as a database for image feature vectors and also as a search engine. The image features can be incorporated into ES index thanks to the elasticsearch-image plugin and LIRE. The latter piece of software includes a set of numeric descriptors based on MPEG-7 edge detection and color histogram algorithms. We extended LIRE with new image features and mixed them with an image-cropping utility, which was possible thanks to OpenCV. We used OpenCV because of its efficient algorithms and also with the hope that in the future we'd be able utilize GPGPU functionality within it. Imagator's user interface was created as a single-page web application in AngularJS. It was used to test a variety of image features, regions of interest and similarity metrics.

The second part of the project was to prepare a production-ready system. This required the development of two additional components: an indexer service and a public API (Figure 4).

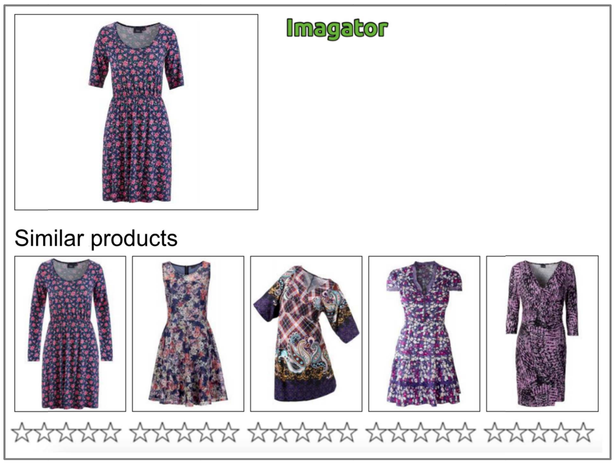

Figure 3. Our test application - Imagator

The indexer service is a subscriber in our internal events queue. It receives all events pertaining to new, deleted and changed offers. We filter out unnecessary categories and offers outside of the Brand Zone and then index the images into our ES index.

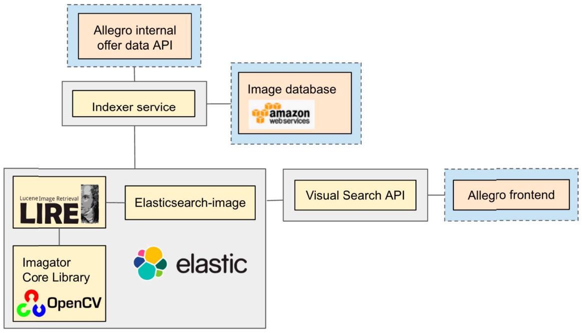

Figure 4. Architecture of our production-ready system.

A public visual recommendation API was needed to interface between the image retrieval backend and recommendation boxes in the Allegro frontend. The results from our visual search API are shown in a separate layer after clicking on a *Show similar* button in a recommendation box (Figure 5).

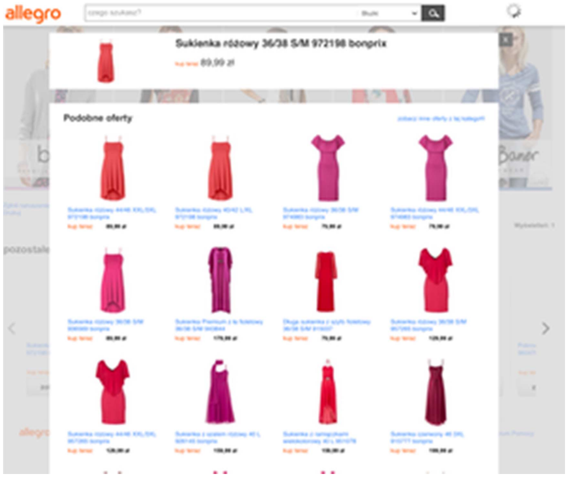

Figure 5. Screenshot showing the system in action within Allegro's production environment.

3.2 Image features and similarity metrics

The LIRE library contains an extensive suite of image features, focusing both on color and pattern extraction. In this work we spotlight the following ones:

- auto color correlogram (ACC) [9]
- edge histogram (EH) [14]
- binary patterns pyramid (BPP) [3][11]
- joint histogram (JH) [11]

Additionally, we implemented a new feature, palette power (PP), as described in [2].

Extracting data from an image is only a first part of the equation. In order to compare images and assess their similarity, we need to compare the results that we get from the above image features. We tried several distance comparison metrics and in the end settled for three that gave the best results: Tanimoto Distance [12], Earth Mover's Distance [13] and Hellinger Distance [7].

4. TESTS

To measure the precision of our image-based search, we conducted both automatic and user-focused tests. The effects of the deployment to production can be quantified thanks to the CTR (click through rate) metric.

4.1 Automatic tests based on textual tags

To assess the difference between descriptors (and also between similarity metrics), we performed automatic tests using textual color tags as a benchmark on a set of about 1300 images from a category containing dresses. We measured the mean average precision (MAP) for positions 1 through 5 in our search listing. Top N selected images similar to the query image are marked as a good choice when they have the same text benchmark parameter. We conducted tests for different sets of text benchmark parameters. We chose among color, pattern or style parameters. However, it's worth mentioning that the results are not very precise and may be too high because of the low quality of textual offer parameters and their low relevance to real visual product descriptions, which was mentioned in section 2.

This automatic test results ranged from 100% for MAP@1 to 60-70% for MAP@5 (Figures 6-8).

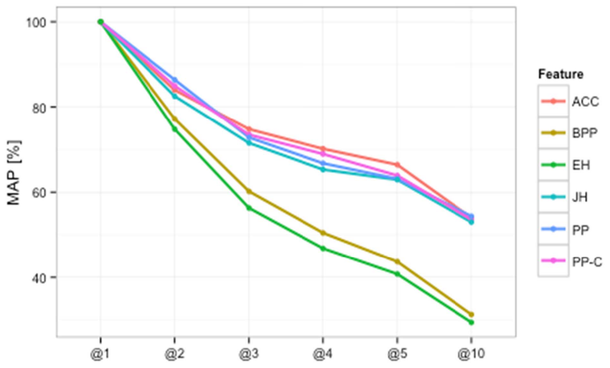

Figure 6. Mean average precisions at positions 1-10 for color parameter values as a text benchmark.

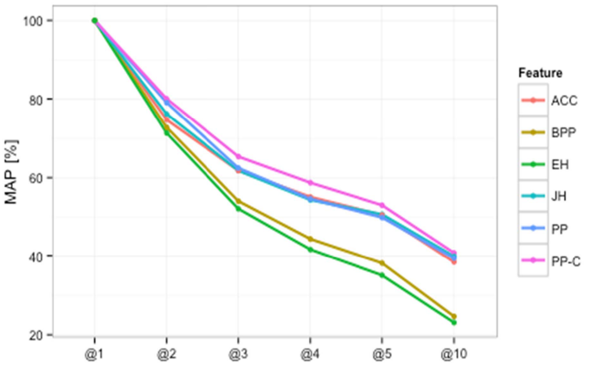

Figure 7. Mean average precisions at positions 1-10 for color and pattern parameter values as a text benchmark.

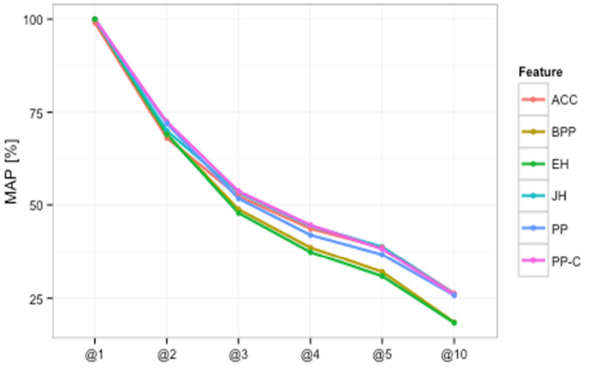

Figure 8. Mean average precisions at positions 1-10 for color, pattern and style parameter values as a text benchmark.

4.2 Subjective user assessments

Imagator shows a view with a large main image and 5 smaller images (Figure 3), deemed as similar by our system (5 images having the highest ranking in ES). We showed this app to about 10 people. We gave the users a simple task to assess the visual similarity of smaller images to the main image using a 5 point scale, where 1 means no correspondence between the main image and the small one and 5 – very similar products in both pictures.

These subjective tests were performed many times - we wanted to gather as much data as possible in order to make decisions on design strategies, i.e. which image features are the most suitable for a particular domain and which segmentation technique is the best. We chose 66 images as main test images. Each of them was used as input in the ES query to find similar images amongst all 1166 indexed images. The main test images were chosen so that their color and pattern parameters must have had similar distribution as the distribution of parameters in all active offers in the dresses category (within the Brand Zone).

Results for different image descriptors settled in the range 2.8 - 3.2 (Table 1, Figure 9).

Many of the users remarked that although the association was not perfect, most of the time there were some similar aspects, e.g. similar tone of color or a comparable pattern. They also said that such recommendations could be very attractive for users who would browse through offers in our service with an intention to buy a roughly similar product.

Table 1. User subjective assessments (UA) for different feature descriptors (-C suffix denotes a central ROI).

Test features	Average UA	UA StdDev	UA Number
1st test phase			
ACC-C users assessed only color similarity	2.82	1.09	990
ACC	3.26	1.27	750
EH-C users assessed only pattern similarity	2.90	1.23	990
ACC-C + EH-C users assessed both characteristics (color and pattern similarity)	2.8	1.04	990
2nd test phase			
JH	3.32	1.05	330
PP	3.08	1.44	465
JH+PP	3.05	1.26	575
BPP+JH	2.41	1.41	655

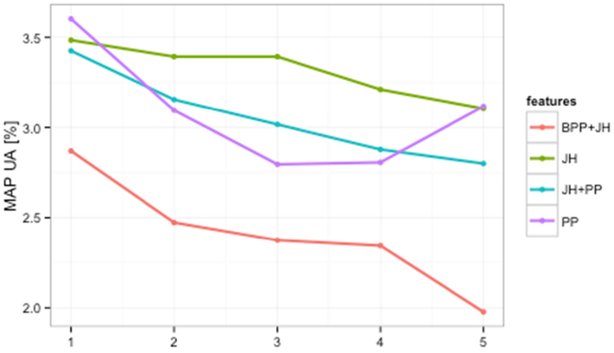

Figure 9. Mean average precision of user assessments (MAP UA) at 1-5 position for 4 feature combinations in subjective tests.

4.3 Production results and discussion

Finally, after analyzing our test results after the 1st phase of tests, we've chosen the best image descriptor (ACC) and similarity metric for Brand Zone images and decided to deploy our solution to operate in a production environment. We integrated it with a recommendation box on the offer description page. Initial A/B tests performed on small user traffic show click through rate results comparable to other techniques based only on textual tags (taking into account the same category).

The subjective results from the 2nd test phase and numbers taken from the production environment show that our product has a great potential to improve, especially when we consider combining the image–based features with textual offer tags.

5. CONCLUSION

Thanks to existing open-source tools it is easy to create an image recommendations system from scratch. The insight of our team members allowed us to fully utilize this opportunity, which resulted in giving Allegro users a new way to browse the service. However, there are still a few challenges ahead of us. We need a way to automatically measure the image quality, which will in turn allow us to appropriately adjust the image matching methods. We are also considering mixing textual descriptions with numeric features extracted from images. Finally, we would like to develop more sophisticated deep learning methods to find visual similarity between our offers.

6. ACKNOWLEDGEMENTS

We would like to sincerely thank Monika Cierzniak (our product owner) and Maria Borowy (our UX designer) for helping us with the deployment to production and with the UX design of Imagator respectively, as well as for all the great ideas and insights.

7. REFERENCES

[1] Allegro online marketplace platform: *http://allegro.pl*.

[2] Bhardwaj, A., Das Sarma, A., Di, W., Hamid, R., Piramuthu, R. and Sundaresan, N. 2013. Palette power: enabling visual search through colors. *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2013), 1321–1329.

[3] Bosch, A., Zisserman, A. and Munoz, X. 2007. Representing shape with a spatial pyramid kernel. *Proceedings of the 6th ACM international conference on Image and video retrieval* (2007), 401–408.

[4] Bradski, G. 2000. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*. 25, 11 (Nov. 2000), 120–126.

[5] Elasticsearch content-based image retrieval plugin: *https://github.com/kzwang/elasticsearch-image*.

[6] Elasticsearch search engine: *https://www.elastic.co*.

[7] Hazewinkel, M. 2002. *Encyclopaedia of mathematics*. Springer-Verlag.

[8] Hsiao, J.-H. and Li, L.-J. 2014. On visual similarity based interactive product recommendation for online shopping. *2014 IEEE International Conference on Image Processing (ICIP)* (Oct. 2014), 3038–3041.

[9] Jing Huang, Kumar, S.R., Mitra, M., Wei-Jing Zhu and Zabih, R. 1997. Image indexing using color correlograms. *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (1997), 762–768.

[10] Jing, Y., Liu, D., Kislyuk, D., Zhai, A., Xu, J., Donahue, J. and Tavel, S. 2015. Visual Search at Pinterest. *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2015), 1889–1898.

[11] Lux, M. 2011. Content based image retrieval with LIRE. *Proceedings of the 19th ACM International Conference on Multimedia* (2011), 735–738.

[12] Rogers, D.J. and Tanimoto, T.T. 1960. A Computer Program for Classifying Plants. *Science*. 132, 3434 (Oct. 1960), 1115–1118.

[13] Rubner, Y., Tomasi, C. and Guibas, L.J. 1998. A metric for distributions with applications to image databases. *Proceedings of the Sixth International Conference on Computer Vision* (1998), 59–66.

[14] Won, C.S.W., Park, D.K.P. and Park, S.-J.P. 2002. Efficient Use of MPEG-7 Edge Histogram Descriptor. *ETRI Journal*. 24, 1 (Feb. 2002), 23–30.