

Searching for Design Examples with Crowdsourcing

Nikita Spirin, Motahhare Eslami, Jie Ding, Pooja Jain, Brian Bailey, Karrie Karahalios

Department of Computer Science, School of Art and Design

University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

{spirin2,eslamim2,jieding2,jain17,bbailey,kkarahal}@illinois.edu

ABSTRACT

Examples are very important in design, but existing tools for design examples search still do not cover many cases. Long tail queries containing subtle and subjective design concepts, like “*calm and quiet*”, “*elegant*”, “*dark background with a hint of color to make it less boring*”, are poorly supported. It is due to the inherent complexity of the task, which so far has been tackled only algorithmically using general image search techniques. We propose an approach for design examples search based on crowdsourcing. Out of many explored crowdsourcing configurations we found that (1) a design need should be represented via three query images, (2) crowd workers should assess the query-specific relevance of candidate examples from a pre-built design collection, and (3) a crowd task should contain verification questions about a design need to control quality. To test the utility of our approach, we compared it with Google Images in a query-by-example mode. Based on feedback from expert designers, the crowd selects more relevant and diverse design examples.

Keywords

AMTurk, crowdsourcing, design search, query-by-example

Categories and Subject Descriptors

H.3.3 [Information Search and Retrieval]: Information filtering, Search process, Selection Process; H.5.3 [Group and Organization Interfaces]: Collaborative computing

1. INTRODUCTION

Examples play a crucial role during the design process. Designers use examples for inspiration, to get familiar with the competitive landscape and to explore the space of design possibilities [1]. Adaptation of existing designs to new cases enables faster prototype development and leads to better design outcomes [2]. In recent years many innovative design search tools were built, e.g. [3, 5] and more.

To understand how existing tools address subtle design search needs, like “*calm and quiet designs showcasing people on a dark background with a hint of color to make it less boring*”, we conducted a test-drive. We restricted our analysis

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Copyright 20XX ACM X-XXXXX-XX-X/XX/XX ...\$15.00.

to Google Images¹ because other tools are research prototypes not accessible to the public. Google Images supports two search modes: using keywords and using an image as a query. We submitted a range of textual and image search queries to Google and analyzed the output.

We found that Google Images (Figure 1, top row) accurately captures color and some typography, but it does not support searches for arbitrary objects, patterns, layouts, and feelings. In this work we propose an approach that augments existing tools and addresses the above shortcomings. Specifically, we contribute by describing an overall pipeline and an optimal configuration for crowdsourcing design search, which enables design examples search for subtle design needs. In an evaluation with expert designers, we compare our approach with Google and show that the crowd reliably selects more relevant and diverse design examples.

2. OUR APPROACH

To define a retrieval system, one should specify a query representation, an index, and a ranking algorithm. We developed our crowdsourcing design search pipeline by iterating on these components. An initial attempt to ask crowd workers to provide web URLs for inspiring design examples matching a textual description didn’t work. Submitted URLs were either irrelevant or pointed to top search results from Google, Bing, and design galleries. Examples from design galleries were the most relevant as these websites feature professionally curated designs. In the second iteration we asked AMT workers to submit URLs to examples matching a textual description specifically from one design gallery. The quality of results improved. However, the URLs were still redundant and biased towards the first few pages of the gallery. We attribute this to the fact that AMT workers had to leave the AMT site to copy URLs back to the web form. To control for these effects, in the third iteration, we crawled examples from the galleries. In the HIT we asked workers to toggle checkboxes for examples matching a textual description without leaving the AMT site. It led to a significant increase in results quality and, therefore, we adopted this index-driven approach. In the fourth iteration we explored multiple query representations. The use of three query images to convey a design need to the crowd was among the best approaches. We picked this representation as it also minimizes the designers’ effort on query formulation compared to other alternatives. To increase a design need understanding by workers, we added two comprehension questions to the HIT. Finally, to minimize cost

¹www.google.com/imghp

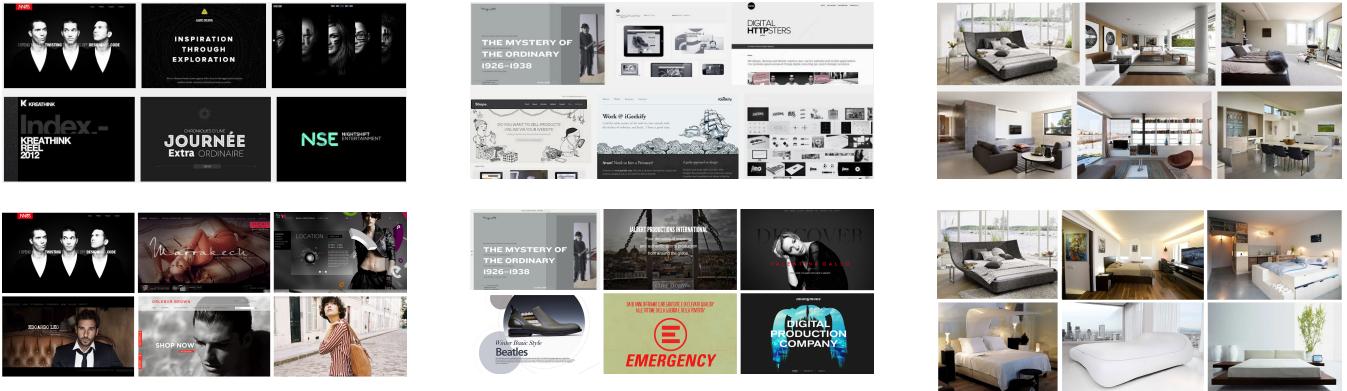


Figure 1: A query image is shown in the upper left corner of each screenshot. *Top:* Results retrieved by Google Images in a query-by-example mode with the scope limited to www.cssdesignawards.com or www.freshome.com. *Bottom:* More relevant and diverse results retrieved via crowdsourcing.

and response time, we introduced the Strike pattern. Rather than asking workers to assess a few examples in a regular HIT, we grouped several HITs in one continuous Strike HIT.

3. EXPERIMENT

We considered real design needs from different design domains. Web design: “*Calm and quiet designs showcasing people on a dark background with a hint of color to make it less boring*”. Typography: “*Typographic designs of a landing page with a solid header font*”. Interior design: “*Light and spacious bedroom with natural lighting and big bed*”. An expert designer found query images to represent each need.

A pool of high quality design examples was built by crawling 2544 examples from www.freshome.com and 961 examples from www.cssdesignawards.com. We used our optimal configuration to post tasks to AMT. The Strike HIT consisted of 6 tabs with 15 examples each. The reward was \$0.11 per HIT. Examples were ranked via majority voting of 9 workers.

For Google Images we limited the scope to only 1 website using an advanced `site:` operator. It guaranteed that both our approach and Google Images performed search over the same set of examples and, hence, the differences were due to the ranking algorithm. Because Google Images allows to specify only one image as query and our approach uses three query images to convey a design need to the crowd, we issued three different queries to Google and built an aggregated ranking by merging the results positionally.

We asked three professional designers to judge the relevance of top-50 examples retrieved by our approach and

Google Images for our experimental design needs. To avoid bias in the evaluation, during the judging process we merged the results together and presented all 100 images without revealing the retrieval method. If an example appeared in both result sets, it only appeared once. Expert judgements were combined using the majority voting scheme.

Google (Figure 1, top row) correctly captures overall content organization and color properties from a query image, but has many irrelevant results. Moreover, inter-result similarity is very high for Google, which makes such results of satisfactory utility for designers. On the contrary, the results retrieved by our approach (Figure 1, bottom row) are relevant and diverse. These are two desired properties for a design search engine as they help designers find inspiring examples and escape from the design fixation problem [4].

The performance of each retrieval method is summarized in the Table 1. Our approach outperforms Google Images across all cut-off points for precision. Remarkably, even for high levels of recall ($P@50$), our approach achieves excellent results. The quality is high for three distinct design domains.

To conclude, in this work we proposed an approach for crowdsourcing design example search and demonstrated its utility. Our approach complements existing algorithmic methods and can be used to handle complex design queries. We plan to conduct a comprehensive evaluation of this idea aiming to uncover new general insights about crowdsourcing.

4. REFERENCES

- [1] S. R. Herring, C.-C. Chang, J. Krantzler, and B. P. Bailey. Getting inspired!: understanding how and why examples are used in creative design practice. CHI ’09.
- [2] W. L. M. Kolodner, J. L. Case-based creative design. In *Proceedings of AAAI on AI and Creativity*, 1993.
- [3] R. Kumar, A. Satyanarayan, C. Torres, M. Lim, S. Ahmad, S. Klemmer, and J. Talton. Webzeitgeist: Design mining the web. In *Proceedings of CHI ’13*.
- [4] A. Purcell and J. Gero. Effects of examples on the results of a design activity. *Knowledge-Based Systems*.
- [5] D. Ritchie, A. A. Kejriwal, and S. R. Klemmer. d.tour: style-based exploration of design example galleries. In *Proceedings of UIST ’11*.

Table 1: Performance of Google Images and our approach for different cut-off points and design queries.

Query Type	Algorithm	P@1	P@5	P@10	P@50
Web	Google	1.0	0.6	0.5	0.34
	Crowd	1.0	1.0	0.9	0.72
Typography	Google	1.0	0.4	0.5	0.42
	Crowd	1.0	1.0	1.0	0.76
Interior	Google	1.0	0.2	0.30	0.16
	Crowd	1.0	1.0	0.80	0.60