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 example search still do not cover many cases. For instance, 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. This is mainly due to the inherent complexity of the task, which so far has been tackled only algorithmically using general image search techniques. We propose a powerful new approach based on crowdsourcing, which complements existing algorithmic approaches and addresses their shortcomings. Out of many explored crowdsourcing configurations we found that (1) a design need should be represented via several query images and (2) AMT crowd workers should assess a query-specific relevance of a candidate example from a pre-built design collection. 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 design examples.

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: Information filtering, Search process, Selection Process

Keywords

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

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 the 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 to Google Images because other tools are research prototypes not accessible to the public. Having analyzed the output for a range of textual and image queries submitted to Google Images, we found that it accurately captures color and some typography (Figure 1, top row), but it does not support searches for arbitrary objects, patterns, layouts, and feelings. [3, 5] push the boundary forward, yet the functionality is still limited due to the reliance on the algorithmic techniques. In this work we propose an approach based on crowdsourcing that complements existing tools and addresses their shortcomings. We contribute by describing an overall pipeline and an optimal configuration for crowdsourcing design search. In an evaluation with expert designers, we compare our approach with Google Images and show that the crowd reliably selects more relevant design examples.

WWW'14 Companion, April 7–11, 2014, Seoul, Korea.
ACM 978-1-4503-2745-9/14/04.
http://dx.doi.org/10.1145/2567948.2577371.

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 an AMT 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 the 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 the 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 speculate that it happened because the AMT workers had to leave an AMT site to copy URLs back to the HIT 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 near 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 the workers, we added two comprehension questions to the HIT but it didn't increase

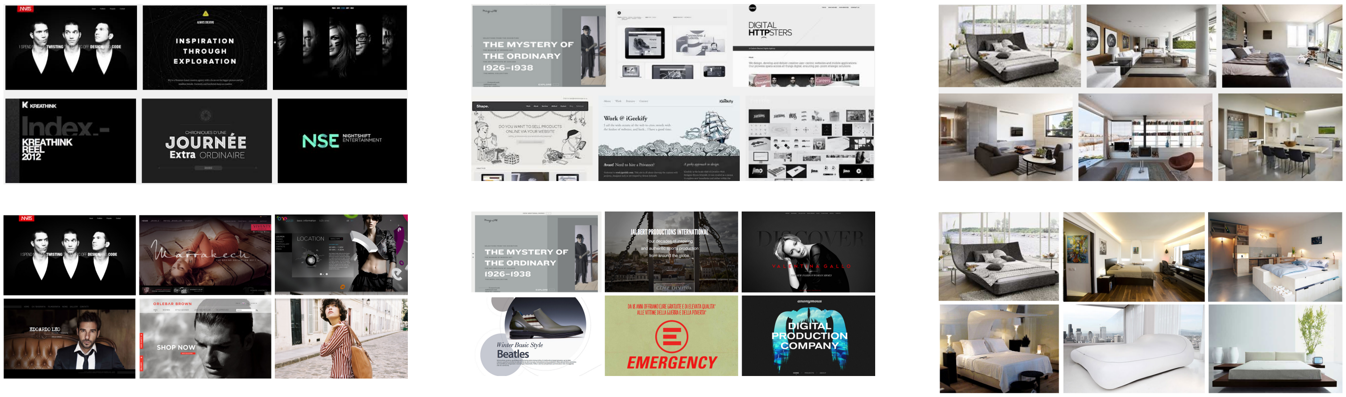

Figure 1: A query image is shown in the upper left corner of each screenshoot. *Top:* Results retrieved by Google Images in a query-by-example mode with the scope limited to *www.cssdesignawards.com* or *www.freshome.com*. *Bottom:* More relevant and diverse results retrieved via crowdsourcing.

the quality of results significantly. Finally, to minimize cost and time, we applied the Strike crowd programming pattern by grouping 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.freshhome.com` and 961 examples from `www.cssdesignawards.com`. We used our optimal configuration to post tasks to AMT. The "strike" HIT had 6 screens of 15 examples. 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, we showed it only 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 minimal 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 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 subtle design queries. We plan to conduct a comprehensive evaluation of this idea aiming to uncover new general insights about crowdsourcing.

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

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

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*.