

Head Crowd: Visual feedback for design

Britta Kalkreuter and David Robb

Abstract

HeadCrowd is a collaboration between the School of Mathematical and Computer Sciences and the School of Textiles and Design at Heriot-Watt University. It investigates how rich web and mobile applications may be employed to provide designers with near instantaneous and highly visual feedback from thousands of potential customers, or crowds. We are exploring the use of state of the art rich media applications to add quantity, speed and statistical accuracy to the study of semiotics, and the use of visuals in fashion as communication.

The project seeks to add to participatory design and market intelligence processes by enabling rapid and iterated co-design cycles between crowds and designers based on visual forms of communication so as to mirror the highly visual nature of fashion design inspiration. Such a scheme shows applications for sustainability in fashion if it can give crowds a concrete sense of ownership of the design process and provide enthusiastic target markets, thereby offering potential to significantly reduce the risks of producing unwanted product.

The paper provides an analysis of prior knowledge before describing the first two stages of the project, in which a pilot browser has been constructed that allows observers to navigate a vocabulary of 500 images which have been ordered into 48 similarity stacks using a mixture of human and crowd sourced sorting techniques. A first test involved the presentation of 20 terms to observers and asking them to choose 3 images from the browser to represent each term. Analysis of the resulting pilot data has given insights into the communicative certainty that a selection of 3 images from a vocabulary of 500 can provide for certain types of terms, and amongst certain groups of testers. It has also prompted deeper analysis of the pilot browser.

To put the communicative value of visual feedback to the test, the current research phase is preparing the reverse experiment of asking a fresh cohort of participants to associate images back to the original terms, and various interfaces are currently being constructed to facilitate the presentation of visual choices from phase 1. The similarity relationship between test images is investigated and visualized before a case is made for comparative experiments of raw selection data and versions of visual summaries in this second research phase, in order to test which way of data presentations best convey the intended visual communication.

Visual feedback, rich web applications, participatory design, sustainability

HeadCrowd: Visual feedback for design

HeadCrowd is a collaboration between the School of Mathematical and Computer Sciences and the School of Textiles and Design at Heriot-Watt University. It will employ rich web and mobile applications to provide designers with near instantaneous and highly visual feedback from thousands of potential customers and collaborators, involving them in the development of products that will have added value to them, thus contributing to sustainable design solutions.

Visual prompts for designers are routinely used in commercial forecasting, and individual ambitions in fashion design are encouraged by a growing number of websites promising co-design, and while web 2.0 and similar and subsequent platforms have opened these new possibilities of participation to designers, forecasters and consumers, the semiotics of visual feedback, and the potential challenges and advantages surrounding its use for creative conversation between different stakeholders, has so far been under explored. Using visual feedback and participatory design as a way to ensure less fast and more sustainable fashion therefore needs to be explored to ensure that the technology goes beyond its potential of just another marketing tool in an ever more competitive market, starting with the exploration of current thinking in the areas of co-design, semiology and visual feedback:

Co-design

In 1980 Toffler posed the term “prosumer” (Toffler 1980) for a consumer who has taken on some responsibilities of work that was previously done for them in the production of a product or service which they simply consumed. Our understanding of Co-design fits well with Toffler’s idea as the consumer in HeadCrowd takes on some of the responsibility for the design by influencing its iteration.

Many of Toffler’s predictions (made from the trends he had observed up to 1980, pre-Internet) have come true. One particularly pertinent to the Head Crowd project is “De-massification of the Media”, which describes how people’s interests are fragmenting into groups and their interests are becoming more specialised. This more individualised consumption of media is an aspect of the current trend in social media. It is hoped that the “Crowd” in the Head Crowd project will consist of interest groups who may become aligned with specific designers. Designers may be able to develop a following among the larger crowd which coalesces around their specific designs.

One aspect often mentioned alongside co-design is mass customisation (Piller et al. 2005). Some of today’s prosumers are catered to by businesses which offer services allowing individuals to specify their own individual requirements. These requirements are usually limited to a set of options but the levels of customisation can be quite sophisticated with Adidas e.g. allowing individualised specification of sports shoe construction and Lego encouraging a user to create their own Lego set with instructions and box graphics.

In contrast, the system envisioned by the project is one where rather than customisation catering to individual specifications, cycles of co-design are to be facilitated with crowds providing input as reaction to designs proposed by designers.

An example of an existing business gathering the opinions of an internet crowd on clothing designs prior to manufacture is the vintage-style clothing online retailer, ModCloth.com. On “Be the Buyer” (Modcloth.com 2012), garments not yet in production are shown and members of the site, who need not necessarily be customers, can give their opinions to influence whether or not they are, in fact, produced. The opinion can be in the form of positive or negative votes and a comment. The comments are in free text but guided towards several specific areas such as colour and “wearability”, with a majority of comments focusing on the former.

It is hoped our proposed system will be taken up by retailers and used to improve on-line promotion of designers and products, generating measurable gains in the marketplace in the short term. However, Sanders & Simons (2009) assert that additional benefits may accrue from co-creation in the form of long term gains in brands and even social gains including improved quality of life and improved ecological sustainability.

A major factor in co-design is the context in which the co-design activity takes place, or the design space. Sanders & Westerlund (2011) describe this as encompassing three aspects: The actual physical space in which the activity happens, the experience and practice of those taking part in the co-design activity along with any current ideas that they are working on, and the domain of all possible solutions to the current design problem. Sanders & Westerlund point out that the first two of these aspects are clearly important in the eventual outcome. In the case of our envisioned system these would equate, in the first case, to the environment provided by the application on terms of communication tools for the designer and the crowd. The second aspect would be some function of the designer and the crowd’s previous design or co-design experience and their expectations. Sanders & Westerlund also make the point that consideration should be given to how the co-design activity will be recorded and how that record will be communicated. It is hoped that the record of each co-design process in our proposed system will add value to an associated final product.

Crowds and Crowd-sourcing

Surowiecki’s “The Wisdom of Crowds” (2004), gives many examples to support his thesis that the collective judgment of a crowd is often superior to that of an expert. He does make the proviso that for the crowd to be relied upon it must be diverse, large, and independent. The independence of each member of the crowd is a property to which Surowiecki assigns particular importance. Examples of when crowds have been seen to fail occur when independence breaks down and individual opinions are influenced by a group mentality such as during stock market bubbles.

Indeed in this regard the aims of the Head Crowd project face a tension. A designer may value the “wise” judgement of the crowd and be expecting added value to accrue in any new design based on crowd feedback. However, another hope of the project is that the crowd will fragment into followings for the different designers. If that occurs, and such an individual following begins to form a group mentality, then the “wisdom” of the crowd could be compromised. If this does happen it is hoped that the value added (and purchases made) within the designer’s following will outweigh any loss of global value in the finished design due to any degradation in the global crowd’s judgement. Indeed it may be sensible for the feedback from the designer’s following to be collected and analysed separately from the global crowd. The designer could receive two separate streams of feedback and make design decisions accordingly.

Semiology and communication

It can be argued that the work of this project seeks to establish a visual language for fast intuitive visual feedback. In that case a brief examination of some principles of linguistics and semiotics is appropriate.

Semiology is usually taken to be the study of signs and symbols. Chandler (2002) defines semiotics as “...the study... of anything which ‘stands for’ something else.” When the web site summarises hundreds of such visual messages, will what the designer reads in the message truly reflect what the crowd was trying to say?

Saussure, in his theory of language, as described by Culler (1976), argued that, in language, signs are an arbitrary combination of signifier and signified. e.g. there is no natural reason for the word, dog, to signify what we recognise as the furry animal that barks. This raises the prospect of an involved and time-consuming language learning process to be gone through before the users and designers can communicate. However, with the abstract image set it is hoped to capitalise on current visual conventions already within the experience of the crowd and the designers to allow communication both to take place initially and to develop.

The work so far in the project has focussed on using abstract images to capture meaning. We have been unable to find any studies involving transmission of meaning using abstract images. However visual communication is already often done with symbols. Signs without words at airports and on our roads are clear evidence that symbolic visual communication works. Otto Neurath (1936) developed a language of pictures to be used in education. Indeed pictographic languages such Japanese use characters originally derived from stylised drawings. Hebecker & Ebbert (2010) have investigated the development and recognisability of free-drawn symbols in response to stimuli terms using a Pictionary-like online game. A communication channel using established signs or emoticons could be a valid component of a visual feedback system.

Image Browsing

Heesch (2008) makes the point that with content based image retrieval a user is required to provide a query. The query can be an example image, a sketch. (This is equally so when providing a text query in a system where images are labelled). However, the user may not always know how to formulate their query. They may not have a representative image to hand or may not have adequate skills or tools to sketch one. Equally any keywords they come up with for use in a label based system may be misleading or too prescriptive. Indeed they may not really know what they are looking for. In these circumstances allowing the user to browse an organised collection of images can allow them to find an image that fits their requirements however vague those requirements were, initially.

Halley (2012) demonstrated that for a large set of monochrome texture images the Self Organising Map (SOM) browser is a good solution and significantly better than two other browsers to which it was compared¹ in terms of time taken for a user to find a given image.

Visuals to words

Taking into account what we already know about visual feedback, HeadCrowd considers design feedback capabilities of images in their own right, but uses the written word in experiments to verify what feedback was intended and picked up by giver and recipient of that message respectively. To minimize the predetermination of visual choice, the project uses a large image set, thus allowing for as free a selection as possible, and asks feedback givers to select three out of 500 visuals for each term described.

The first technical phase of the research concentrated on shaping prototype interfaces that allow fast and intuitive navigation of big numbers of visuals to facilitate selection by crowds; this essentially meant developing algorithms and protocols which automate the organization of large rich data sets. Easy navigation of these data sets was key to obtaining results that are relatively unaffected by fatigue, and our current work has made use of research on perceptually relevant image browsing by Halley (2012). That work showed that the Self Organising Map (SOM) browser was more effective than other ways of organising images to allow retrieval of a given image.

Visuals

In order to arrive at an appropriately large and varied image set, some 1800 images tagged as abstract were screen scraped from Flickr.com account holders who make their images available with a Creative Commons licence. Although only images tagged as “abstract” were used, a large number of miss-tagged images had to be discarded manually as they included images of people, full conventional depictions of objects and writing. Duplicates were excluded by a process of screening the image

¹The two other methods were “Rapid-fire”, a method proposed by Wittenburg *et al.* (1998), and a 3D MDS-based browser, suggested by Rogowitz *et al.* (1998).

pixel data, which meant sorting the images by average red, green and blue pixel level and then displaying the images in order side-by-side in screens of multiple images. (One duplicate was found). Images which were very similar in composition to others were eliminated by hand so as to try to prevent a number of similar images occupying spaces in the set thus reducing the breadth of the available image “vocabulary”.

Limitations of pilot visual data set

A number of difficulties of this modus operandi were identified through analysis of the first pilot selection data: A small number of near duplicates found in the final image highlighted the limitations of the pixel level elimination exercise, and can also be attributed in part to human fatigue when faced with the scale of image selection needed for this experiment (1800 images narrowed down to 500). Furthermore, they indicated to the team that the number of Flickr accounts (or other appropriate sources) searched for images needed to be greater for future data sets.

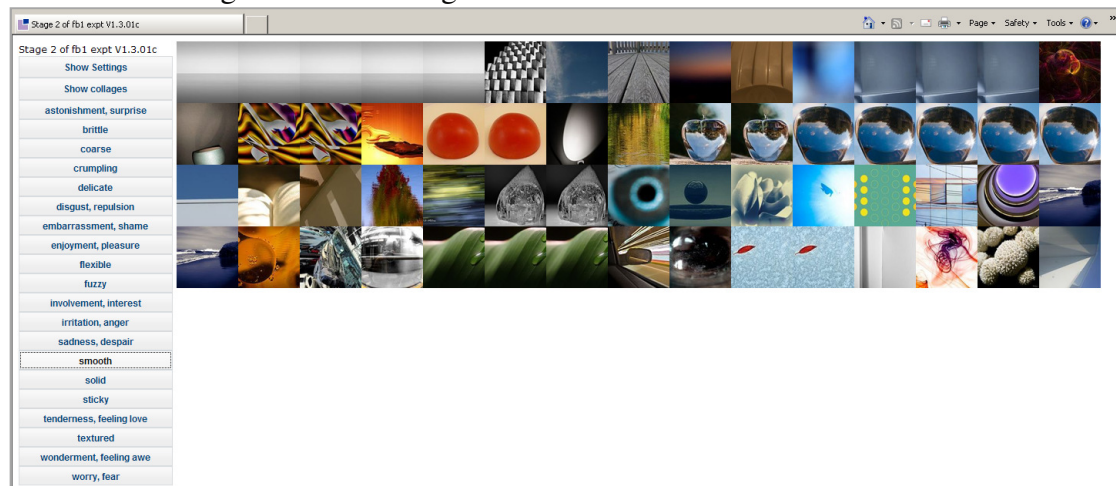


Figure 1: Screenshot of image selection for 'smooth'.

An example of near duplicates are the small and big glass apples on the screen in figure 1 which shows the full image selection by 20 observers for the term “smooth”; typical for the appearance of near duplicates in the pilot data, observers here clearly demonstrate their capability of identifying distinct images as similar despite their obvious and stark differences in scale, whereas an automated image sorter working on pixel levels could not. Since they are therefore treated as entirely unrelated images as far as electronic tagging is concerned, their obvious similarity will have to be considered as a limitation when evaluating statistics on frequency of identical image choices.

This observation highlights the demand to use more sophisticated automated image sorting in future experiments.

Another realisation of the pilot data analysis was that a significant number of images made their way into the final 500-strong image set despite not actually following the

project's definition of abstract as they show people, full conventional depictions of objects or include writing or accepted symbols.

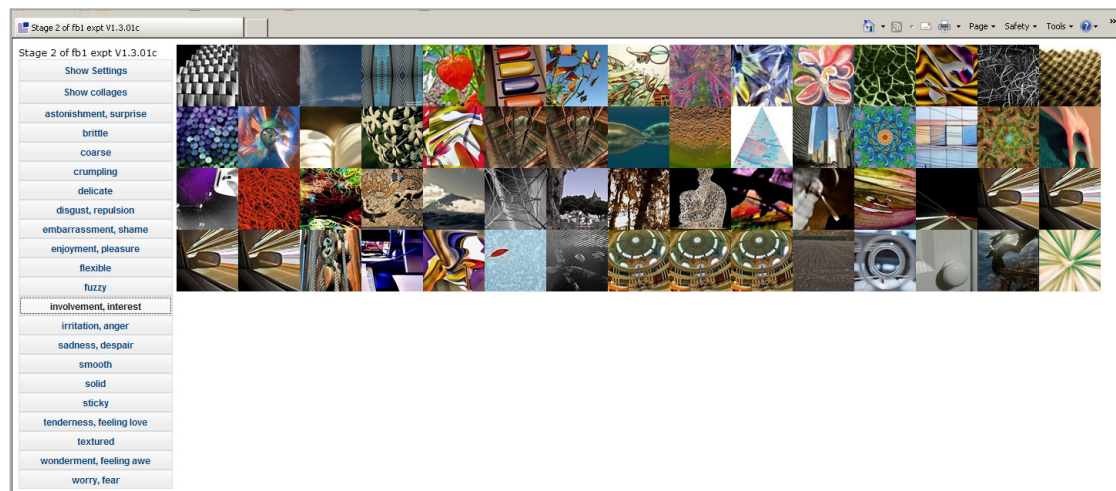


Figure 2: Screenshot of image selection for 'involvement, interest'.

Again, this deviation from the originally intended selection criteria can partly be explained by human fatigue in the face of having to scrutinise such a large set of relatively small images, but it also calls for a tighter set of criteria for future image sets and further investigation of whether the margin for human error in privately tagged image collection makes them suitable for use in a controlled experiment.

However, while clearly a potential limitation of the current image set's usefulness, we were able to use the existence of these non-abstract images in the pilot study to evaluate the varying efficacy of different image types for certain purposes, and certain groups when communicating visually: Selections for the term "involvement and interest" by the 20 pilot testers (see figure 2) for example demonstrate that descriptive images rather than truly abstract ones are popular when observers are asked to associate images with emotional terms, as opposed to choices for material terms such as "smooth" (figure 1), where instances of semi-figurative images are notably lower. This trend holds firm over all 10 emotional and 10 material terms in the visuals to word pilot.

Organising the images into manageable interfaces

While HeadCrowd seeks to employ computer technology whenever appropriate, it recognizes earlier research that indicates the importance of a human basis for such an undertaking: The Self organizing Map (SOM) browser, as developed by Halley [2012], is effective because it is based on collected judgements of humans about which images are similar to each other. For a small number of images this can be done by having observers freely group the images into however many similarity groups they wish, and 20 observers organized the first 100 images of this project in this way. Each time a given image is grouped with another given image, that specific pairing of

images is given a similarity score of 1 before the score is added up and then divided by the number of opportunities there were for that pairing to have occurred. The resulting similarity matrix for the initial 100 is 100 rows by 100 columns and can be used to categorise the images based on their perceived similarity.

To eliminate the dangers of human fatigue while sorting, this nuclear, human sorted image set was then enlarged to a visual “vocabulary” of the desired size of 500 by a crowd sourcing method devised by Halley [2012] and also employed by Padilla et al [2012]. It presents a “bootstrap” SOM of the 100 hand sorted images to observers sourced through Amazon Mechanical Turk (MTurk) who use this bootstrap SOM to provide similarity judgements on small packets of the further images (termed the augmentation images). Each MTurk observer is shown 20 of the augmentation images, 1 at a time, and asked to select bootstrap images they judge to be similar to the query image.

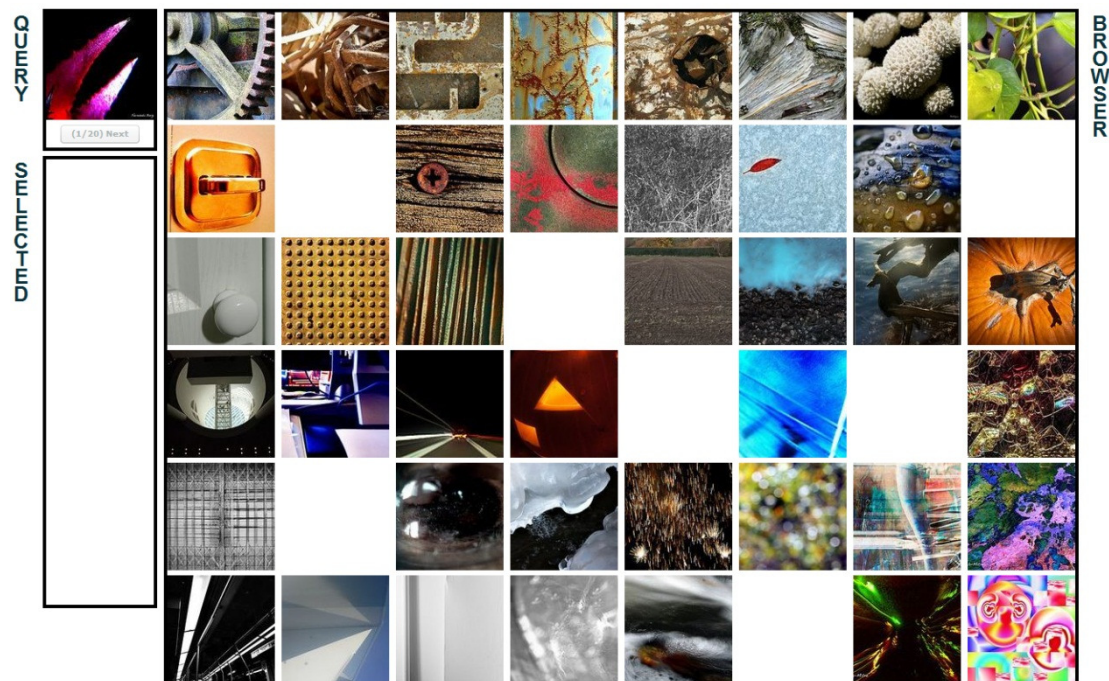


Figure 3: A screen seen by an MTurk observer during the augmentation process.

Figure 3 shows a screen seen by an MTurk observer during the augmentation process, where they are shown the query image (top left) and have to choose between 2 and 4 images from the bootstrap SOM as being those most similar to the query image. The data from the observed likenesses between the 100 bootstrap images and the 400 augmentation images was used to calculate the augmented 500x500 similarity matrix. It is this larger similarity matrix on which the pilot study SOM browser is based. For the purpose of presenting this SOM on an iPad to the observers in the pilot study, an 8 x 6 grid was chosen. Figure 7 shows the full, 8 x 6 stack SOM browser.

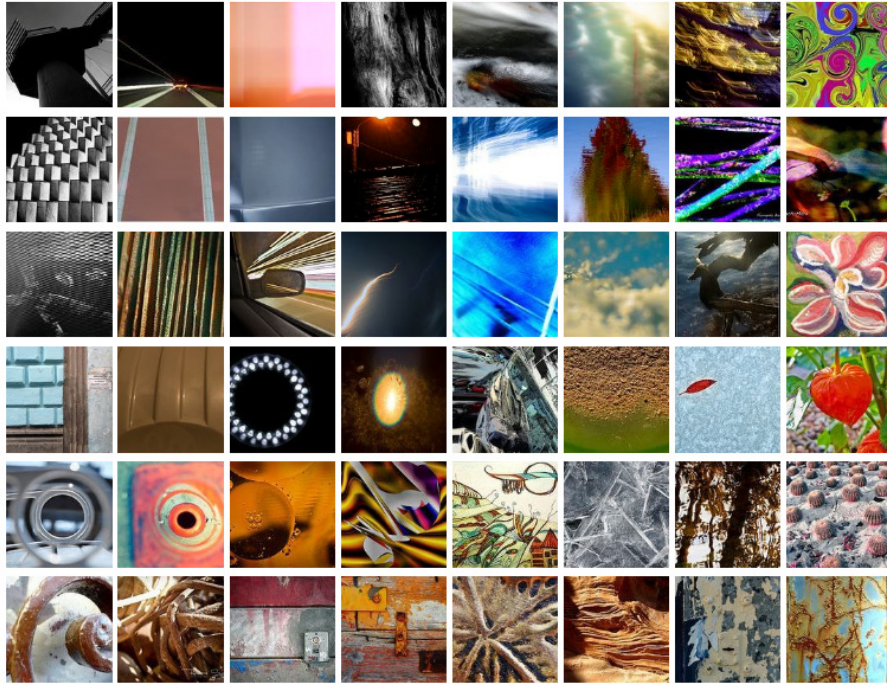


Figure 4: The full, 8 x 6 stack, SOM browser on iPad

The image on the top of a particular stack in the SOM browser is the image in the stack which is nearest the centroid position of the perceptual similarity space represented by all the images in that stack. Once an observer clicks on one of the images on this top level, that centroid image, together with all images that were associated as similar (and have therefore been assigned to its stack) are shown on a screen for the observer to choose from. (See Figure 5).

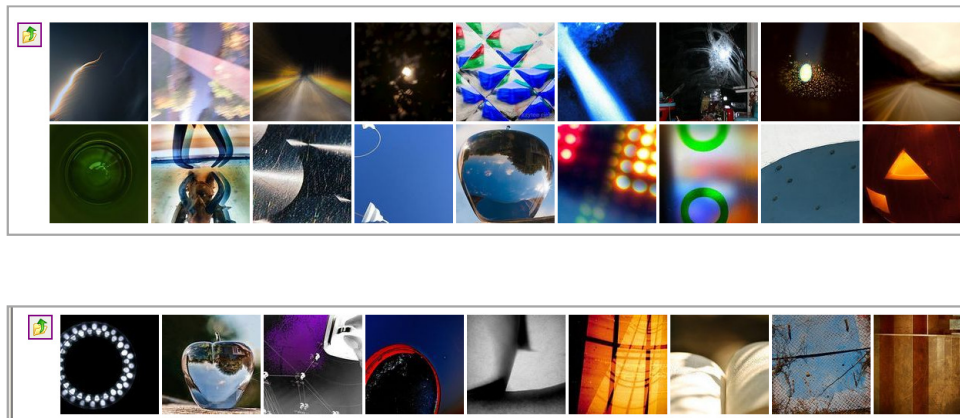


Figure 5: The stacks beneath two adjacent images (flash and ring of light) just left of the centre of the SOM browser. Note two glass apples in separate, but adjacent stacks

A detailed analysis of all pilot image choices will need to consider their position on the SOM browser as a possible determining factor for selection or non-selection by an observer. Already it emerges that although the 48 top level (6x8 iPad interface)

images only make up less than 1/10 of the entire image set, their representation in final image choices is nearer 1/7, suggesting their prominent position has made them more likely to be chosen by observers.

However, image position in the stacks and on the SOM browser interface is, of course, far from random as it correlates to the complex interrelationship of similarities that was assigned to images by initial hand sorting and MTurk augmentation:

Visualising similarities between images in stacks, and images in adjacent stacks through 3D MDS visualisation of the similarity space

A highly navigable way to observe how images had been assigned to similarity stacks by hand sorting and MTurk augmentation is to visualise the perceptual similarity information for the 500 images in a 3D multidimensional scaling (MDS) view. The view can be used to “swim” back and forth amongst the images plotted in 3D similarity space. (See Figure 6), relating back to the SOM data obtained through both sorting exercises.

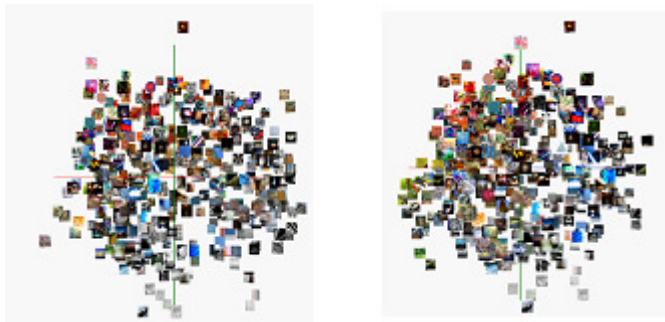


Figure 6: Screenshots of the distribution of the 500 images plotted in their perceptual similarity space in a 3D MDS view. The right hand screen shot was taken after the visualisation was rotated about 90 degrees.

The 3 dimensions depicted in the 3D MDS view describe over 80% of the variability in the similarity data and can also be viewed in clusters. (See figure 7)

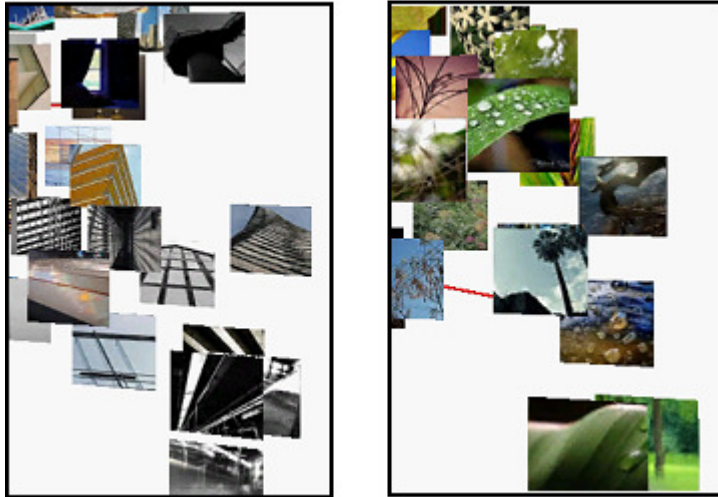


Figure 7: Two clusters of images from the 3D MDS visualisation. The cluster of images on the left is very much on a structural theme while the cluster of images on the right is on a natural theme. Much work remains to analyse the relationship between position and selection of images.

Communicating visually

With the aim of discovering whether meaning can be reliably communicated between participants of different backgrounds using a fixed set of images we are conducting a selection experiment in two phases:

Phase 1 Visual to word pilot

Phase 1 of the pilot has already been completed and involved the presentation of 20 terms to 20 observers, asking them to choose 3 images from the SOM browser to represent each term. The terms comprised of 10 descriptive words sourced from non-experts as being descriptive of fabrics or products [Methven et al 2011] and 10 emotional terms sourced from the “Geneva Emotional Wheel” [Scherer 2005].

20 Observers (10 male, 10 female, 10 designers, 10 non-designers from mixed (European and Asian) backgrounds in the 20-30 year age bracket) were shown the 10 descriptive terms followed by the 10 emotional terms in a random order and chose 3 different images from the SOM to represent each given term. The experiment was presented to the observers on an iPad, and their responses were recorded on a central server, making data collection and analysis easy, for almost any size of future sample. Analysis concentrated on criteria and research questions based on semiotic and sociological questions surrounding the image set and selection behaviour of the distinct observer groups in the pilot:

Quantitative evaluation of the pilot data, namely considering the frequency with which a particular visual was chosen by more than one observer to describe the same

term, was identified as a key measure for assessing how fixed (and strong) the communication value of an image is amongst respondents. Some interesting general trends emerged as well as some specific ones when filtering image choices by specific groups, and for specific terms:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	phrase	group	range	no of singles	%singles	no of double	%doubles	no of triples	%triples	no of quadruple	%quadruples	no of quintuples	sixtuplets	septuplets
2	brittle	All	18-488	38	63.33	5	16.67	2	10	0	0.00	0	1	0
3	coarse	All	5-486	25	41.67	10	33.33	1	5	3	20.00	0	0	0
4	crumpling	All	20-495	30	50.00	8	26.67	2	10	2	13.33	0	0	0
5	delicate	All	12-494	39	65.00	5	16.67	2	10	0	0.00	1	0	0
6	flexible	All	8-490	38	63.33	4	13.33	3	15	0	0.00	1	0	0
7	fuzzy	All	9-483	35	58.33	8	26.67	3	15	0	0.00	0	0	0
8	smooth	All	17-499	32	53.33	6	20.00	2	10	0	0.00	2	0	0
9	solid	All	5-495	24	40.00	4	13.33	4	20	4	26.67	0	0	0
10	sticky	All	52-496	34	56.67	7	23.33	1	5	1	6.67	1	0	0
11	textured	All	5-495	37	61.67	7	23.33	3	15	0	0.00	0	0	0
12	astonish	All	1-494	37	61.67	4	13.33	1	5	3	20.00	0	0	0
13	surprise	All	3-490	40	66.67	8	26.67	0	0	1	6.67	0	0	0
14	disgust	All	11-492	52	86.67	4	13.33	0	0	0	0.00	0	0	0
15	embarr/	All	2-492	38	63.33	8	26.67	2	10	0	0.00	0	0	0
16	enjoym/	All	1-495	51	85.00	1	3.33	1	5	1	6.67	0	0	0
17	interest/	All	4-484	38	63.33	6	20.00	2	10	1	6.67	0	0	0
18	irritation/	All	12-440	41	68.33	8	26.67	1	5	0	0.00	0	0	0
19	anger	All	8-487	33	55.00	5	16.67	1	5	1	6.67	0	0	0
20	sadness/d	All	4-494	44	73.33	4	13.33	1	5	0	0.00	1	0	0
21	espair	All	16-474	43	71.67	4	13.33	3	15	0	0.00	0	0	0
22	tendern/l	All		37.45	62.42	5.8	19.33	1.75	8.75	0.85	5.67	0.3	0.05	0

Figure 8: Screenshot of results processing spreadsheet.

Taking the average from all 20 observers on all 20 terms, 62.42 % of choices were singular, meaning that 62.42% of the images that made it into the 60-strong selection for each term had not been associated with this given term by more than one observer.

Across all 20 terms and 20 observers, just 19.33% of images were chosen by two individuals as being descriptive of the same term, 8.75% of images were agreed on by three observers, and 5.67% were selected 4 times as being illustrative of the same meaning.

Regarding this result according to the Saussurian concept of signifier and signified means that observers in the majority disagreed about the visual signifier for a given signified; in other words, the result indicates the image set’s relatively low certainty of communicative value and may be seen to confirm that a lack of codification in visuals renders them a less suitable or certainly less unambiguous tool for imparting intelligible feedback than language, and should therefore be avoided. However, HeadCrowd is aimed at developing tools for a design discipline that works on highly visual inspiration rather than relying on textual description, and looking at the pilot data in more detail and according to certain search criteria, allows an interesting qualification of the rather negative first impression on communicative value of visual feedback:

When considering all of the 20 observers’ choices separately by type of term, it emerged that singular choices, seen above as an indicator for low communicative

certainty of an image, accounted on average for 69.50% of selections for the emotional terms but for just 55.33% of images when observers had been asked to describe textural terms.

While fashion undoubtedly consists of a whole lot more than cloth stitched into shape, it means that the terms that are arguably more immediately meaningful to fashion design found more agreement across all 20 observers on which image might represent them. In fact it seems somewhat remarkable that almost half of the image choices for textural terms were agreed on by at least two of 20 individual observers, each faced with choosing just 3 per term from a bank of 500. Despite this outcome, emotive terms cannot be discarded in this study of visual feedback for design as they are frequently present in aforementioned forecasting and mood boards alike. The conclusion we must draw from this part of the pilot evaluation is however that particular care must be taken when eliciting and evaluating feedback on these more ephemeral aspects of design.

Even less agreement on the most appropriate visual signifier of a given term can be detected when looking at the pilot data from certain social groups' (or distinct crowds') point of view:

Sorting the choices by gender, males top the list of disagreeing by choosing on average 79.17% of images just once while females display just 74.33% of singular choices. Looking at the pilot data according to professional background, designers and non-designers interestingly do not on average display a less or more unambiguous attitude to using visual signifiers (as might have been expected given different work practices as regards visuals) as both groups share an identical 77.17% of singular image choices.

Looking at the pilot data in detail rather than by averages confirms some of the above described trends, e.g. when "solid" scores the lowest percentage for singular choices (i.e. the highest level of agreed communication) across all 20 observers with just 40%, while "involvement/ interest" scores 100% of individual choices, meaning total disagreement of communication value. Detailed scrutiny of the pilot results also shows some startling deviations from the norm as females score just 50% of singular choices on "tenderness and love", showing an unusually high level of choice overlap for an emotional term, while 93.33% of designers' choices for "delicate" are not matched by any fellow professionals, though material terms generally generated more agreement. Minute scrutiny of the results will continue to identify all factors that may have had a significant influence on choices. Amongst these is the way in which the 500 images were presented on the iPad interface:

Is the number of images in the full set proportionate to the choices allowed each observer, i.e. can observers easily handle 500 images?

Do observers of a certain background (female, male, designer, non-designer) favour certain types of images to communicate specific things? A thorough analysis of mood boards and forecasting images will be conducted to help answer this query.

With the help of eye-tracking equipment, can we determine whether the position of images on the iPad SOM browser has an effect on its frequency of selection, e.g. top left hand corner rather than centre of page.

As indicated above, does the position of an image on the SOM browser have an effect on its selection, i.e. are top level stack images selected more frequently than those buried deep in stacks, thus requiring the observer to click through to them.

Are the images chosen more than once distinct in terms of position within the iPad interface or within certain stacks.

As this semiotically motivated analysis of the pilot continues, HeadCrowd's premise to use visuals as suggestive parts of a cumulative whole (the visual aggregation) is being researched in terms of a computer's capability of calculating and representing what such a collage of image choices might look like, when informed by a set of collected human judgements about image similarity. A pilot aggregation of the 60 image selections per term according to a raft of popularity and similarity factors has now been designed to be used to facilitate the reverse experiment of eliciting term selections in response to images, or as Saussure would say, selecting a signified to the signifiers presented:

Visualisations of the responses in Phase 1

The responses were visualised in two ways: All the chosen images for a given term were displayed simply in order of image ID number. Images that were chosen more than once were displayed multiple times to indicate this. (See Figures 1 and 2)

The images chosen for each term were displayed in 3D similarity space with the size (area) of each image being increased proportionately to signify its popularity. (See Figure 9 below).

Visual summaries of the image selections

As the eventual volume of image feedback selections by an internet crowd is envisioned to be very large, unmanageable in fact for visual feedback, a way of summarising the selections was needed as an alternative to simply listing the images. To this end a cluster analysis was carried out on the chosen images based on their similarity data. The responses were divided into 10 clusters. The image nearest to a cluster's centroid in the similarity space was chosen to represent that cluster in a summary collage which will then consist of just 10 images of size depending on each

cluster's popularity. The position of the images is determined by their position in the similarity space, projected into two dimensions. (See figure 11)

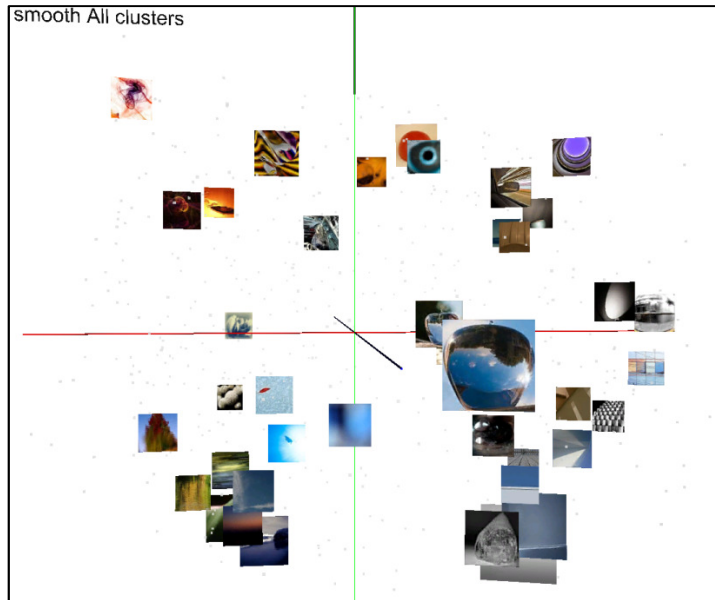


Figure 9: screenshot from 3D visualisation of the image selection for 'smooth'. This shows the full selection (all clusters) with size (area) varying with popularity. Note two glass apples.

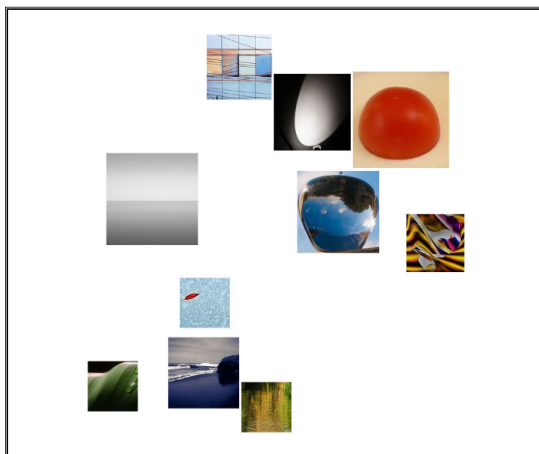


Figure 10: Summary collage of image selection for 'smooth'. Each image represents a cluster of images. The size (area) of an image represents the population of the cluster. See all image selections for 'smooth' in figure 1

The next phase of the pilot will be a comparative study showing the raw (60 image) selections and their automatised summary collages to a fresh set of observers in random

order, asking them to indicate which of the 20 terms they think the given view represents. It will provide firmer answers as to how well meaning can be conveyed using mass selections from large image set. Showing raw selection data and aggregated summary separately to the same observers will indicate whether summary collages are more or less effective than full sets of image selections at conveying meaning, with potentially wide ranging consequences for the usability of crowd-sourced visual feedback for design.

References

- Cohen, S. S. (1988). *Practical Statistics*. London, Edward Arnold (Hodder and Stoughton).
- Chen, C., Gagaudakis, G. & Rosin, P. (2000) Similarity-Based Image Browsing, *Proceedings of Int'l conference intelligent information processing*, Beijing, 22 August 2000. pp 206–213.
- Clarke, A. D. F., Dong, X. & Chantler, M. J. (2012) Does Free-sorting Provide a Good Estimate of Visual Similarity, *Predicting Perceptions: Proceedings of the 3rd International Conference on Appearance*, Edinburgh, UK, Lulu Press, Edinburgh, UK, pp 17-20.
- Cox, T. F. & Cox, M. A. A. (2001) *Multidimensional Scaling*, Chapman & Hall/CRC Boca Raton, Florida, USA.
- Craft, B. & Cairns, P., (2006). Using Sketching to Aid the Collaborative Design of Information Visualisation Software-A Case Study, In *Human Work Interaction Design: Designing for Human Work*, Springer Boston, Volume (221), p. 103--122.
- Culler, J. (1976) *Saussure*, Fontana
- Delort, J.-Y., Arunasalam, B. & Paris, C. (2011) Automatic Moderation of Online Discussion Sites. *International Journal of Electronic Commerce*, 15 (3), pp. 9 - 13.
- Estelles-Arolas, E. and F. Gonzalez-Ladron-de-Guevara (2012). "Towards an integrated crowdsourcing definition", *Journal of Information Science* 38(2): 189-200.
- Everitt B. (1974) *Cluster Analysis*, Heinnemann.
- Google (2012), "Google help: Search by Image: How it Works", <http://support.google.com/images/bin/answer.py?hl=en&answer=1325808> [accessed 22 Jun 2012]

- Google (2011), "A fall spring-clean", Google Official Blog September 2, 2011
<http://googleblog.blogspot.co.uk/2011/09/fall-spring-clean.html> [26 Jun 2012]
- Graham, D. J.; Friedenberg, J. D.; Rockmore, D. N.; Field, D. J. (2010) Mapping the similarity space of paintings: Image statistics and visual perception. *Visual Cognition*, 18 (4), 559-573.
- Guiraud, P. (1971). *Semiology*. London, Routledge.
- Halley, F. (2012). "Perceptually Relevant Browsing Environments for Large Texture Databases", PhD thesis, Edinburgh, Heriot Watt University, Edinburgh.
- Heesch, D. (2008). "A survey of browsing models for content based image retrieval." *Multimedia Tools and Applications* 40(2), pp261-284.
- Hebecker, R., Ebbert, C. (2010) Creation and Validation of Symbols with Purposeful Games and Online Survey, *Proceedings of the 1st DESIRE Network Conference on Creativity and Innovation in Design*, Aarhus, Denmark, Desire Network, Lancaster, UK: Aarhus, Denmark pp. 112-120.
- He, D. C. & Wang, L. (1990) "Texture unit, texture spectrum, and texture analysis". *IEEE Transactions on Geoscience and Remote Sensing* 28.
- Kerminen, P.; Gabbouj, M. (1999) "Image Retrieval Based on Color Matching", FINSIG'99, Oulu, pp 89 – 93.
- Kerminen, P.; Gabbouj, M. (2000) "Visual Goodness Evaluation of Color-based Retrieval Processes", In EUSIPCO 2000, Tampere, Finland, pp 2153-2156.
- Kerminen, P.; Tantt, J. T.; Gabbouj, M. (2003) Utilization of luminance information on color-based image retrieval, Nordic MATLAB Conference 2003, Comsol A/S, pp 252 – 256.
- Methven, T.S., Orzechowski, P. M., Chantler, M. J., Baurley, S. & Atkinson, D. (2011). "A Comparison of Crowd-Sourcing vs. Traditional Techniques for Deriving Consumer Terms". Digital Engagement '11, November 15 – 17, 2011, Newcastle, UK [online at <http://de2011.computing.dundee.ac.uk/wp-content/uploads/2011/10/A-Comparison-of-Crowd-Sourcing-vs.-Traditional-Techniques-for-Deriving-Consumer-Terms.pdf>, accessed 20 June 2012].
- Modcloth.com (2012), "Be the Buyer"
http://www.modcloth.com/storefront/products/be_the_buyer
[accessed 26 Jun 2012]
- Neurath, O. (1936) *International Picture Language*. Kegan Paul, Trench & Trubner Co., London.

- Padilla, S., Robb, D., Halley, F. & Chantler, M. J. (2012) Browsing Abstract Art by Appearance., *Predicting Perceptions: Proceedings of the 3rd International Conference on Appearance*. Lulu Press, Edinburgh UK, pp. 100-103. ISBN 978-1-4716-6869-2
- Piller, F., Schubert, P., Koch, M., & Moslein, K. (2005). "Overcoming mass confusion: Collaborative customer co-design in online communities." *Journal Of Computer-Mediated Communication*, 10(4).
- Rogowitz, B. E.; Frese, T.; Smith, J. R.; Bouman, C. A. & Kalin, E. (1998) Perceptual image similarity experiments., *SPIE Conference on Human Vision and Electronic Imaging*, pp 576-590.
- Sanders, E. B.-N. & Simons, G. (2009). "A Social Vision for Value Co-creation in Design." Open Source Business Resource(Value Co-Creation).
- Sanders, E. B.-N. & Westerlund, B. (2011) Experiencing, Exploring, and Experimenting in and with Co-Design Spaces. In Nordic Design Research Conference, Helsinki., 2011.
- Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information*, 44(4), pp.693-727.