Do deep features retrieve X?: Project update

Julieta Martinez
`julm@cs.ubc.ca`

November 2015

1 Progress update

I have abandoned the idea of using images of different sizes, which also makes tiling trivial. This should not be a big problem, since it is a common step to take a central crop of the images before computing deep features [5]: therefore, this image representation actually reflects more accurately the input used to compute the deep features. I have compiled the "Faces in things" dataset, making each image of size 128 × 128; such that the smallest dimension is fully preserved. To download the images I used tweepy [1], and I made a few python scripts to merge the images in a big single tile[1] to minimize the number of http requests.

I have found that, unfortunately, d3js is not widely used to display images (as these are usually shown on a `canvas`, which unlike `svg` does not have an internal DOM structure of its own). The only example I was able to find with d3 and images is from a visualization of the front row of 2013's Fashion Week [2]. I have adapted some pieces of this code to explore a tile of images with a Cartesian fisheye distortion[2]. Informally, I asked two people in the vision lab to take a look at the interface, and the feedback has been that it looks "slick" and "cool", but it is also "dizzying". I have noticed that the Fashion Week visualization has smoothed transitions, so I'll try to implement that in an effort to reduce the dizzying effect of my visualization. Alternatively, I will implement a static, size-encoded similar to that of Chen *et al.* [6], and add on top a focus+context exploration tool.

All the code that I have written so far can be accessed on `https://github.com/jltmtz/deepviz`.

[1] The tile can be accessed on `http://jltmtz.github.io/deepviz/imgs/facespics_128/bigtile.jpg`

[2] A demo of my progress so far can be accessed on `http://jltmtz.github.io/deepviz/`. It runs smoother on Google chrome.

2 Previous work

The goal of this work is to provide a tool to quickly inspect the output of an image retrieval system; however, a secondary goals is to allow users to explore the whole dataset (in the context of an arrangement relative to a query), to allow users to explore different arrangements resulting from using several images as queries (and thus, through several examples, get an idea of which visual concepts are retrieved). For the first part, we review work that focuses on image tiling for large visual datasets, and for the second part we focus on work that addresses focus+context exploration in large image collections.

2.1 Image tiling

There is a large amount of previous work focused displaying multiple images with different types of arrangement and tiling. Google images[3] shows a set of relevance-sorted images in vertically-stacked horizontal layouts, controlling for height and preserving the image aspect ratio. While it is not clear whether this visualization is optimal, it is probably the one that people are most familiar with, and has been subsequently been adopted by competing image search engines such as Bing and Flickr. In 2009, Google introduced Image Swirl [3], which clustered similar images in stacks along the z-dimension, and allowed users to expand clusters for closer inspection. However, the service was only available as beta, and seems to have been discontinued.

Recently, Brivo *et al.* [4] proposed using a Voronoi-based paritition of the space to show multiple images in space-filling manner. After a force-directed layout is computed from the query to the rest of the dataset, a Voronoi diagram is used to resize and crop the images relative to their distance to the focus image. The authors further propose a method to smoothly change the diagram when new images are added or removed from the collection. The advantage of this method is that it is easy to implement, and achieves space-filling in an unconventional way. The disadvantage is that the cropping of Voronoi cells may confuse some users, who are mostly used to deal with rectangular crops. While the method is novel, the authors did not make a user study comparing to a standard rectangular arrangement and cropping.

Wang *et al.* [7] propose a tiling of images around a central query that uses size to encode importance. The query image is shown in the center with maximum size, and the most relevant results are shown around the image in a spiral layout, halving the size with each full wrap around the query. The advantage of this method is that it conveys the idea that retrieval importance is non-linear and rapidly decreases as one explores images further down in the ranking. The downside is that there is not an easy way to allow users to focus on images far from the query, and that the spiral layouts makes it hard to compare the relative ranking of any two retrieved images with respect to the query.

[3] https://images.google.com/

2.2 Focus+context in image collections

The second task that we want to support is allowing users to explore the image collection to use new images as queries, with the output of a previous query as context. Previous work on focus+context in images is rather scarce. The only work that we are aware of is that of Chen *et al.* [6] contrast and study sliding (where neighbours are simply pushed away) and expanding (where neighbours are re-arranged according to a Voronoi diagram to better use the space) approaches to focus+context exploration. They found that users prefer the expanding (space-filling) approach in terms of “ease to use”, “efficiency” and “fun”. This suggests that we should prefer visualizations that densely populate the space in the focus+context exploration part of our tool.

A somewhat related work (although definitely closest in terms of implementation) in this vein is the NYT Fashion Week front row visualization by Michael Bostock [2]. The online system lets users explore several front shots of models wearing designer clothes. The images are aligned horizontally, and a 1-dimensional fisheye zoom focused on the user’s mouse is used to expand the images for closer focus+context inspection. This is a simpler problem compared to ours, as (a) there is already a precompiled visual structure and similarity in the pictures – notice that a central slit is shown in the cropped images, and this slit manages to give an idea of the attire −, and (b) fisheye zooming is only done in one dimension, as horizontal scrolling is used to show the work of different designers. In contrast, our work has to deal with changing (computed on-the-fly) visual similarities, and also an arrangement in 2 dimensions.

References

[1] https://github.com/tweepy/tweepy.

[2] http://www.nytimes.com/newsgraphics/2013/09/13/fashion-week-editors-picks/.

[3] https://googleblog.blogspot.ca/2009/11/explore-images-with-google-image-swirl.html.

[4] Paolo Brivio, Marco Tarini, and Paolo Cignoni. Browsing large image datasets through voronoi diagrams. *Visualization and Computer Graphics, IEEE Transactions on*, 16(6):1261–1270, 2010.

[5] Ken Chatfield, Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. Return of the devil in the details: Delving deep into convolutional nets. *arXiv preprint arXiv:1405.3531*, 2014.

[6] Jiajian Chen, Yan Xu, Greg Turk, and John Stasko. Easyzoom: Zoom-in-context views for exploring large collections of images (Technical report). 2013.

[7] Chaoli Wang, John P Reese, Huan Zhang, Jun Tao, Yi Gu, Jun Ma, and Robert J Nemiroff. Similarity-based visualization of large image collections. *Information Visualization*, 2013.