# Image Ranking with Density Trees for Google Maps

Jared M Johnson
Google, Inc.

Sema Berkiten
Google, Inc.

Top cluster, *image courtesy Victor R. Ruiz.* 2nd cluster, *image courtesy stantontcady* 3rd cluster, *image courtesy Pierre Blaché* 4th cluster, *image courtesy Jed Scattergood*

5th cluster, *image courtesy Ron Cogswell* 6th cluster, *image courtesy Daniel Lobo* 7th cluster, *image courtesy kafka4prez* 8th cluster, *image courtesy Mobilus In Mobili*

**Figure 1: The National Mall (Washington, D.C.) Flickr* photos most similar to the top images chosen by our algorithm.**

## ABSTRACT

We propose an unsupervised learning technique for image ranking of photos contributed by Google Maps users. A density tree is built for each point-of-interest (POI), such as The National Mall or the Louvre. This tree is used to construct clusters, which are then ranked based on size and quality. We choose a representative image for each cluster, resulting in a ranked set of high-quality, diverse, and relevant images for each POI. We validated our algorithm in a side-by-side preference study.

**ACM Reference Format:**
Jared M Johnson and Sema Berkiten. 2020. Image Ranking with Density Trees for Google Maps. In *Special Interest Group on Computer Graphics and Interactive Techniques Conference Talks (SIGGRAPH '20 Talks), August 17, 2020.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3388767.3407353

## 1 INTRODUCTION

Large scale online platforms, such as Google Maps, receive vast numbers of user-generated content on a daily basis. As the amount of contributions increases, the challenging problem of organizing and presenting this content in a useful way within limited space becomes more prominent. Given a set of user-contributed images for a single POI, our goal is to create a *representative* and *high-quality* roster that is visually and semantically *diverse*.

As with most user-generated content, some contributions fall short; low quality images or those that are not relevant to the POI are present in the input dataset. In this work, we leverage the fact that content relevant to the subject will agglomerate at a greater rate than irrelevant imagery, which can be de-emphasized. Our method is robust to the large percentages of irrelevant or spam imagery that users contribute in practice, even for POIs with few images.

We propose using Gaussian Kernel Density Estimation (KDE) to construct a tree in a pixel-based embedding space for all the images of a POI. Next, locally dense image clusters are obtained using graph cuts on this density tree. We rank these clusters based on their size and image quality. Finally, we select images from each cluster based on their quality and the distance to the root of their subtree. For the most popular POIs in the world, user contributions number in the millions, our method represents these huge sets of imagery beautifully, concisely, and without redundancy (Figure 1), making optimal use of the usually limited space reserved for images.

## 2 RELATED WORK

We break down the problem of ranking images into orthogonal steps: (1) grouping similar images, (2) ordering the groups, and (3) finding an ideal representative image for each group. K-means clustering [MacQueen 1967] is one of the most common unsupervised classification algorithms and our first approach at grouping


*SIGGRAPH '20 Talks, August 17, 2020, Virtual Event, USA*

ACM ISBN 978-1-4503-1234-5/17/07.
https://doi.org/10.1145/3388767.3407353

[0]*All photos in this figure are licensed under a Creative Commons Attribution 2.0 Generic (CC-BY2.0). Accessed 10 Feb 2020, [Fli 2004].

images, but we found it too difficult to adaptively define $k$ and too dependent on initialization conditions. We also investigated utilizing DBSCAN [Ling 1972], a more-predictable spatial clustering algorithm that does not require a pre-defined $k$. Unfortunately, the inconsistent semantic meaning of distances in our embedding space precluded consistent results across many different kinds of POIs.

Li et al. [2016] created an algorithm for initializing $k$-Means more effectively and predictably, utilizing density estimation. Wu et al. [2017] showed superior clustering results based on density trees. These investigations lead to our image grouping method based-on density tree graph-cuts. Our approach is formulated specifically to the problem of image ranking, combining (1) an optimal image clustering step with, (2) ranking clusters, and (3) ranking images within clusters.

## 3 METHOD

The foundation of our method is the density result from KDE, calculated from Euclidean distance in a proprietary high-dimensional image embedding space (any image embedding space can be used e.g. [Frome et al. 2013]). This space describes images semantically at a high level in relative terms: Semantically similar images have low distance to one another. The density calculation is done for each image as a normalized summation of the Gaussian Kernel function of the Euclidean distance between each image and all other $N$ images:

$$D(x) = \frac{1}{\sigma N \sqrt{2\pi}} \sum_{i=1}^{N} e^{-0.5 \frac{d(x,x_i)^2}{\sigma^2}} \quad (1)$$

where $\sigma$ is the kernel bandwidth, $d(x, x_i)$ is the Euclidean distance, and $D(x)$ is the density at image $x$ in our embedding space.

### 3.1 The Density Tree

Once each image's density is known, we build a tree over these density values, with each image's parent node being defined as the closest image with a higher density. Consequently, the root node is the image with the highest KDE result and the leaves are all local density minima. Without forming the imagery into clusters, this tree already gives us a lot of information. We know which images have similar content to other images and which images are likely outliers. However, to create a diverse and representative set, we need to further divide the density tree.

### 3.2 Graph Cuts

Distant edges are good candidates for cuts. However, distances cannot be compared throughout our embeddings space; similar distances have different semantic similarities throughout the space. In order to adaptively account for these differences, we frame the problem of detecting tree discontinuities as a graph-cut problem (see examples in supplemental material). We find the edges with the highest cost and cut these edges that are greater than some minimum distance. The primary cost function is the difference between (1) the sum of the edge distance between a node and its parent and all descendant distances to a node's parent and (2) the sum of all descendant distances to that node. Starting from the deepest nodes and ascending, we cut all high-cost edges, generating numerous density sub-trees.

### 3.3 Cluster Ordering

We treat these pruned density sub-trees as clusters of semantically-similar images, which simplifies the dimensionality of our ranking problem: Which semantic content describes the POI the best? Which clusters are the most iconic semantically? The main signal that comes directly from the graph pruning step is the number of images in each cluster. Clusters with many images are frequently photographed by our users and probably the most likely semantics to be imagined when a given POI is thought-of, a good measure of iconicness.

### 3.4 Representative Image Selection

Once clusters have been ordered by their size, we choose an image to represent that cluster. The root of the sub-tree/cluster is a highly-representative image, however, it usually is not the most aesthetically pleasing image within the cluster. Our objective is to select the most aesthetically beautiful image with semantics that closely match the sub-tree root image.

We sort imagery within a cluster utilizing a model that takes in pixels and image metadata and produces a single score measuring a combination of image quality and image usefulness (e.g. a beautiful selfie is of low usefulness on Google Maps). Images in this list are filtered out if they are above median distance from the root node, giving us an adaptive way to guarantee some degree of semantic similarity to the root node. The top image is used as a representative image for its subtree.

## 4 EVALUATION

We validate our ranked results in a side-by-side evaluation with the top five images for 797 POIs were shown to 485 raters. Our results were compared against images ranked by their quality and diversified based on their distances to each other in our embedding space using a greedy approach. Given the name of the place and two sets of images, we asked the raters which side they prefer while considering the *relevance*, *quality*, and *diversity* of the image sets. The base and the experiments sides are flipped randomly before showing them to the raters and three raters evaluated each question. Answers for each question were aggregated using majority voting. 43% of questions did not have any consensus answer. Among the answers with consensus, raters preferred our results 62% of the time, they preferred the base side 21% of the time, and raters found both sides the same 17% of the time.

## REFERENCES

2004. Flickr. http://www.flickr.com

Andrea Frome, Greg S Corrado, Jon Shlens, Samy Bengio, Jeff Dean, Marc' Aurelio Ranzato, and Tomas Mikolov. 2013. DeViSE: A Deep Visual-Semantic Embedding Model. In *Advances in Neural Information Processing Systems 26*. Curran Associates, Inc., 2121–2129.

Fengfu Li, Hong Qiao, and Bo Zhang. 2016. Effective Deterministic Initialization for $k$-Means-Like Methods via Local Density Peaks Searching. *ArXiv* abs/1611.06777 (2016).

R. F. Ling. 1972. On the theory and construction of $k$-clusters. *Comput. J.* 15, 4 (01 1972), 326–332. https://doi.org/10.1093/comjnl/15.4.326

J. MacQueen. 1967. Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*. University of California Press, Berkeley, Calif., 281–297.

Xiaozhu Wu. 2017. SCMDOT: Spatial Clustering with Multiple Density-Ordered Tree. *International Journal of Geo-Information* (06 2017).