Calculating the prominence and isolation of every mountain in the world

Andrew Kirmse
ark@alum.mit.edu

Jonathan de Ferranti
jonathan@viewfinderpanoramas.org

October 31, 2017

Abstract

A global-scale calculation identifies peaks in a digital elevation model (DEM) and computes their isolation and topographic prominence. A new DEM is presented that covers the entire globe at 90 meter resolution with no substantial voids or artifacts. All peaks with at least 1 kilometer of isolation are found, and the closest higher ground is identified. For prominence, all peaks with at least 30 meters are found, and the key saddle is identified. The prominence algorithm uses results from Morse-Smale topology to run in parallel on standard, freely-available elevation data. Thirteen previously unknown "ultra-prominent" mountains with at least 1500 meters of prominence are listed.

I. Introduction

The three most important objective measures of a mountain are its elevation, isolation, and prominence. Elevation is a fundamental property of the Earth's surface that measures the radial distance of a point on the surface above an equipotential surface known as the geoid. Isolation and prominence are derived values that characterize the tendency of a mountain to dominate its surroundings. Isolation is the distance along the surface between a point and the nearest point with higher elevation, known as the *isolation limit point* (ILP). Mountains with high isolation values are of interest to climbers because of their remoteness. Although many major mountains have high isolation, even a small coral atoll several meters high in the middle of the ocean can have an isolation of thousands of kilometers. The isolation of Mt. Everest is undefined, since there is no higher point on the Earth.

Prominence, also called re-ascent or rise, can be defined in multiple equivalent ways:

1. The minimum vertical distance one must descend from a point in order to reach a higher point.

2. The difference between the elevation of a point, and the elevation of the lowest contour line that contains it and no higher point.

When using the first definition, one may imagine walking downhill from a peak A in order to reach a higher peak B. The minimum elevation one reaches will be at a saddle (inflection) point, because there are two paths leading uphill away from the point to A and B, while by the definition, since one must descend to this point before re-ascending, there can be no other adjacent higher ground. This point is known as the *key saddle* for peak A, and the prominence of A is equal to the peak elevation minus its key saddle's elevation. It can be shown that no two peaks share the same key saddle.

The highest point of any land mass, such as an island, effectively has the ocean as its key saddle, and thus its prominence is equal to its elevation. In particular, this is true of Mt. Everest.

1.1 Previous work

Modern attempts to classify peaks are commonly traced to Munro (1891), who attempted to identify "distinct mountains" among those over 3,000 feet high in Scotland. While topographic prominence is now the most common measure used to distinguish separate peaks in mountaineering, other definitions of the identity and extent of a mountain are possible. For example, Tomko and Purves (2010) measured the "apparent prominence" of a peak by analyzing the height of a peak in a photograph above the horizon line. Llobera (2001) defined prominence as the percentage of locations within a given radius that lie below an individual's location. Whereas we consider all local maxima as potential peaks, Podobnikar (2012) filtered maxima based on elevation, isolation, convexity, and other criteria to automatically identify mountaineering peaks and classify the shapes of their summits. Fisher et al. (2004) and Deng and Wilson (2008) examined DEMs at multiple resolutions to come up with fuzzy measures of "peakness." All of these approaches contain extra free parameters (for example, the location of the observer or weighting parameters at each resolution) that make them potentially applicable in disparate applications, but lack the simplicity of a single, objective metric.

The concept of topographic prominence was used since at least the early 20th century as a way to determine which peaks should be considered independent mountains, and which are mere sub-peaks (Jurgalski, 2009). For example, in Colorado in the 1960s, a "saddle drop" of 300 feet (91m) was required for a peak to qualify as a separate "fourteener" over 14,000 feet (4267m) high. Early measurements could be done by survey, or by examining topographic maps. Fry (1987) was the first to apply the term "prominence." From the mid-1990s to the early 2000s, Aaron Maizlish, Andy Martin, Edward Earl, and others began assembling lists of the most prominent peaks in the world, using a threshold of 1500 meters, and the most prominent peaks in the United States, using a threshold of 2000 feet (610m) (Martin, 1994). They laboriously examined topographic and aeronautical maps by hand, tracing contours to find the key saddles, which in some cases could be hundreds of kilometers away from the location of the corresponding peak.

In 2003, the first digital elevation model (DEM) with near global coverage was made available to the public (US Geological Survey, 2004). One of us (de Ferranti) developed an algorithm to calculate prominence from this data based on the second definition given above. This algorithm starts at a given peak, then iteratively lowers a target elevation. At each step, the contour line at the target elevation is determined. The lowest contour that contains no higher peak is at the elevation of the key saddle. de Ferranti and Maizlish used this algorithm to fill out their list of the peaks with 1500 meters of prominence. These peaks, numbering just under 1500, became known as the *ultras*, and are primary targets for prominence-minded mountaineers.

The present work is most directly based on algorithms developed by Edward Earl starting in 1998, and embodied in an interactive program called WinProm (Earl, 2016 and Helman, 2005). The first version of WinProm implemented an algorithm that Earl called *discrete local prominent region* (DLPR). To picture the operation of DLPR, imagine pouring a different color of paint on each peak and letting it flow downhill. Key saddles appear at the high points along the boundaries between colors. Earl extended the algorithm to splice neighboring tiles together to analyze larger regions. He used WinProm to find the peaks of California and Nevada with at least 2000 feet of prominence.

Correct identification of saddles proved problematic in DLPR, and in 2001, Earl and Metzler developed a new approach based on Morse-Smale topology (Helman, 2005). In this approach, saddles are identified first, and then lines of steepest ascent are taken until peaks are reached. This generates what is known as the unstable Morse-Smale complex of the DEM (Čomić et al., 2005). Earl was working on a global prominence analysis using this approach at the time of his death in 2015 in a mountaineering accident.

In parallel, Kirk (2016) and associates examined topographic maps for the entire United States by hand between 2004 and 2016. They identified approximately 140,000 peaks with at least 300 feet (91m) of prominence. Beginning in 2013, he augmented his

analysis with a Fortran program that generated contours, similar to de Ferranti's program. This identified an additional 1400 peaks.

Early peak databases computed the distance from each peak to the closest higher peak in the database, which is an overestimate of the true isolation value, and obviously depends on the database's coverage. Maizlish and Slayden (2004) determined the 42 most isolated mountains in the world by manually reviewing maps. Since then, the isolation and ILPs of selected major mountains have been determined by examining topographic maps, or in a few cases, by computer program for a very small set of major peaks. Lists of peaks with 25 miles (40 km) of isolation were assembled for various subregions of the United States, 100 miles (161 km) for Alaska, and 1000 kilometers for the world.

This paper describes a new computer analysis that identifies the isolation and prominence of every peak in the world, down to the resolution possible with the current best available elevation data. While calculating isolation is essentially by brute force, the prominence calculation uses concepts first introduced in WinProm. The results are the first comprehensive databases of peak isolation and prominence values for the planet.

II. Methods

2.1 Input data

DEMs are often delivered as files with 1 square degree of coverage. Each square (or *tile*) is further divided into a uniform square grid of elevation samples. The resolution of the tile is characterized by the length of a sample at the equator. For example, in 3-arcsecond data, each sample corresponds to an area 3 arcseconds on a side at the equator, or approximately 90 meters. For 1-arcsecond data, the nominal spacing is 30 meters, and for 1/3-arcsecond data, it is 10 meters. As latitude increases, each sample covers a smaller and smaller length in the longitudinal direction, proportional to the cosine of the latitude.

For much of the world, the best available elevation model is derived from the Shuttle Radar Topography Mission (SRTM), which has 3-arcsecond resolution. While this data represented a major leap forward when it was published, it has three shortcomings for use in terrain analysis. First and most seriously, 0.2% of the coverage area consists of voids where no data was collected, especially in mountainous areas that are of particular interest (see Figure 1). Second, coverage extends only to 60 degrees North latitude and 56 degrees South latitude. Third, the data is a surface model, meaning that it measures the topmost elevation at a point, including the tops of vegetation and buildings. The elevation of a mountain is considered to be the highest natural point of ground, excluding these other features.

From 2005 to 2012, one of us built an augmented version of the SRTM data that filled essentially all voids, and extended coverage to the entire globe (de Ferranti, 2016). The data for Antarctica is from the Radarsat Antarctic Mapping Project (Liu et al., 2001), resampled from its original 200-meter resolution. Greenland was covered by digitizing topographic maps, combined with a resampled 100-meter DEM from the Geoscience Laser Altimeter on the Ice, Cloud, and Land Elevation Satellite (DiMarzio, 2007). The northern parts of Scandinavia and Russia were covered by digitizing Russian topographic maps. Many voids were filled with samples from the Advanced Spaceborne Thermal Emission and Reflection Radiometer global DEM (ASTER GDEM), though a review of its properties showed that it has too many artifacts near water bodies, clouds and high mountains to make it suitable as the primary database for our analysis (de Ferranti, 2011).

In remaining SRTM voids, a complete DEM was generated by the following method. First, the SRTM data was converted to contours using the standard marching squares algorithm. Next, contour lines were extracted from either digital or scanned paper topographic maps, and matched to the edges of the voids. Any obvious artifacts in the SRTM data were found at this stage by visual inspection, and replaced by data from the topographic maps. Finally, the merged contours were re-digitized to a DEM, preserving the original SRTM elevation values outside the areas with voids or artifacts. Although single-sample voids still exist in the merged global result, the extent of larger

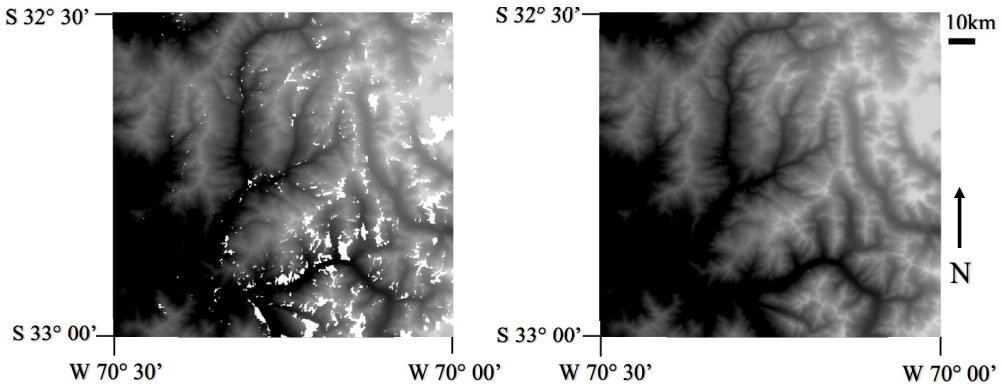

Figure 1: (a) A typical Shuttle Radar Topography Mission scene with many voids (white pixels) in mountainous areas. (b) The same area with voids filled by data from Advanced Spaceborne Thermal Emission and Reflection Radiometer global digital elevation model and topographic maps.

voids is so small that only in three cases are data samples completely surrounded by void samples, and even the largest of these cases is only two data samples in size.

In North America, higher resolution data is available through the National Elevation Dataset (NED, see Gesch et al., 2002). This DEM was assembled from a variety of sources at different resolutions, and processed to match edges and remove most artifacts. In the contiguous 48 states and Hawaii, 1/3-arcsecond data is available. One arcsecond data covers the entire United States, Canada, and Mexico. Using multiple data sets with different resolutions increased the complexity of the calculation, and introduced some small errors along their shared borders, as will be seen. Despite this, we felt that it was important to use the best available data, especially over the U.S., which is the largest region for which an exhaustive list of low-prominence peaks exists (Kirk, 2016). This served as a useful validation during development.

SRTM data represents elevations as integer meters, while NED data stores elevations as floating point meters. The absolute vertical error in SRTM elevations is roughly 6 to 9 meters at 90% (Rodriguez et al., 2005), and for NED data it varies from 1.5 meters root-mean-squared (RMS) in the conterminous U.S., to 3.6 meters RMS in Canada and 6.7 meters RMS in Mexico (Gesch et al., 2014). Error ranges are not incorporated into the present analysis.

Many terrain algorithms require examining areas larger than one square degree. To make this easier, the DEMs are typically processed to make them seamless, meaning that they repeat the exact same values along one or more edges. SRTM tiles are 1201 by 1201 samples in size, with one sample of overlap along each edge. NED tiles are 3612 by 3612 samples for 1-arcsecond data, and 10812 by 10812 samples for 1/3-arcsecond data, with six samples of overlap along each edge. Unfortunately, none of these data sets turns out to be truly seamless in practice. The de Ferranti DEM did not always copy void-filling values to neighboring tiles, so its edges don't always agree. Within the NED data, there are often slight but significant differences in floating point elevation values in the overlap region. To guarantee that the DEM is seamless regardless of the source data, anytime a tile T is loaded into memory, we also load its eastern and southern neighbors, and copy their leftmost column and topmost row into T's rightmost column and bottommost row.

2.2 Isolation

Calculating isolation is a straightforward process. First, all peaks within the DEM are found, and then a brute force search is made to find the closest higher sample to each peak. This point is the ILP, and the isolation value is the distance between the peak and the ILP. A description of this search illustrates the challenges of working with DEM data, and sets the stage for the more complex prominence calculation to follow.

Any sample that is at least as high as its 8 neighbors is potentially a peak. If a peak consists of multiple samples at the same elevation value (i.e. it has a flat top), we want to identify only a single sample to represent the entire peak. To do this, we initialize a separate array of integers, initially zero, of the same dimensions as the tile. This array is called the *domain map*. When a peak sample is found, we fill the domain map with ones where corresponding connected samples in the DEM have the same elevation as the peak. As we continue to scan the DEM for peaks, we skip any sample where the corresponding value in the domain map is one. Any samples with a special elevation value indicating a void are also skipped.

Edge and corner samples have only 5 and 3 neighbors within their tiles, but for the purposes of computing isolation, they can also be considered potential peaks. For efficiency in the peak-finding step, we do not look at any neighboring tiles. If an edge or corner sample turns out to have a higher neighbour in another tile, it will be found at a later stage, and the false peak will be filtered out because of its zero isolation value.

Because DEMs represent the average elevation values in each of their samples, they tend to underestimate summit elevations, sometimes significantly. It's desirable to use the most accurate known summit elevations, as incorrectly identifying one summit as higher than another nearby will result in their isolation values being switched. The Web site Peakbagger.com has a database of approximately 60,000 summit elevations, many of which are known to the nearest foot from accurate surveys. As each tile was read in from disk, the elevations of any peaks from Peakbagger intersecting the tile were written to the corresponding sample. In the usual case where Peakbagger's value was higher, this would achieve the desired effect of bumping up the summit elevation to the more accurate value. When Peakbagger had a lower value, this operation would have almost no effect, since neighboring samples would remain higher, ultimately leaving the computed isolation value nearly unchanged. No solution to this problem has yet been implemented. It would be possible to "raze" the summit area down to the Peakbagger elevation value, but it is not clear where this razing operation would stop in order to avoid erasing other nearby peaks.

Most of Peakbagger's surveyed elevations are in feet, rather than meters. To avoid losing precision and possibly misidentifying two peaks as equally high when one is known to be a foot or two higher, all DEM elevations were converted to feet as they were loaded, and the analyses operated on elevation values in feet.

2.2.1 Finding the closest higher ground. Centered on each peak, we check concentric rectangles of increasing size, looking for a sample higher than the peak. In order for these rectangles to encompass all points at a given distance from the peak, they must be taller than they are wide, compensating for the fact that lines of longitude get closer with increasing latitude. The aspect ratio of the rectangle is the cosine of the peak's latitude.

If higher ground is found in the tile, it will be either inside or outside a rectangle's inscribed ellipse that represents the locus of points equidistant from the peak. If the higher ground is inside the ellipse, then the fact that the circumscribed rectangle was already searched means that there is no closer higher ground in the tile. If the higher ground is outside the ellipse, then we must still search one more, larger rectangle to check for closer higher ground. Giving consecutive rectangles a ratio of $\sqrt{2}$ in linear dimensions guarantees that the larger rectangle includes every point as close to the peak as any point from the smaller rectangle.

Of course, the peak's closest higher ground may be in an adjacent tile, or even in a tile very far away. If

we found higher ground in the peak's tile, then we have a maximum distance that we can use to constrain the search in neighboring tiles. If not, then all samples in neighboring tiles must be checked. We compute a spherical cap of the Earth at a given maximum distance, then search all tiles that intersect the cap. The maximum distance continues to increase until higher ground is found, or the entire Earth has been searched.

Whereas the Earth is close enough to Euclidean within a tile that we may use a planar distance formula, as we start to search neighboring tiles, we must use a more accurate but much slower spherical formula:

$$a = \sin^2\left(\frac{\phi_2 - \phi_1}{2}\right) + \sin^2\left(\frac{\lambda_2 - \lambda_1}{2}\right)\cos(\lambda_1)\cos(\lambda_2)$$

$$\text{distance} = R\tan^{-1}\left(\frac{\sqrt{a}}{\sqrt{1-a}}\right)$$

where R is the radius of the Earth, λ_1 and λ_2 are the two longitude values, and ϕ_1 and ϕ_2 are the two latitude values. Once we have the closest higher ground, we can further refine the isolation value by using a more accurate formula that treats the Earth as an ellipsoid. This makes a meaningful difference only for the very largest isolation values.

$$d\lambda = (\lambda_2 - \lambda_1)/2, d\phi = (\phi_2 - \phi_1)/2$$

$$\Lambda = (\lambda_2 + \lambda_1)/2$$

$$s = \sin^2 d\phi \cdot \cos^2 d\lambda + \cos^2 \Lambda \cdot \sin^2 d\lambda$$

$$c = \cos^2 d\phi \cdot \cos^2 d\lambda + \sin^2 \Lambda \cdot \sin^2 d\lambda$$

$$w = \tan^{-1}(\sqrt{s}/\sqrt{c})$$

$$r = \sqrt{sc}/w$$

$$\text{distance} = 2aw\left(1 + f\frac{3r-1}{2c}\sin^2\Lambda \cdot \cos^2 d\phi - f\frac{3r+1}{2s}sin^2 d\phi \cdot \cos^2\Lambda\right)$$

where a and f are the equatorial radius and flattening of the WGS84 ellipsoid respectively (Meeus, 1991).

In order to speed up the computation, a least-recently-used cache of tiles is kept in memory to minimize the number of times the same tile is read from disk. Processing tiles by increasing latitude, and increasing longitude within each row of latitude, improves the hit rate of the cache. As long as the cache is correctly locked, tiles can be processed in parallel by a pool of threads as large as the number of available CPUs.

2.3 Prominence

The strategy for computing prominence is first to convert each DEM tile to a tree data structure where peaks and saddles are nodes. This structure is called a *divide tree* because the edges in the tree represent watershed boundaries. We can merge divide trees from adjacent tiles into larger continent-sized trees by introducing auxiliary nodes called *runoffs*. It is relatively straightforward to compute prominence directly once these large divide trees are built. Finally, as an optimization, we must demonstrate how to prune a divide tree so that even an entire continent's tree can fit in a computer's memory. Figure 2 gives a schematic overview of the algorithm.

2.3.1 Building the divide tree. The DEM's unstable Morse-Smale complex is constructed by starting at saddles, and following two lines of steepest ascent until they reach peaks. Accordingly, given a tile, we must first identify all of the peaks and saddles. As in the isolation calculation, we construct an auxiliary domain map of the same size as the tile, and initialize it with zeros.

We scan the samples in the tile, flood-filling at each one in order to identify all of the contiguous samples with the same elevation. We then assemble a collection of all samples touching the flat region, known as the *boundary*. This boundary may be as small as 8 samples, or many millions of samples in size in low-lying river deltas.

If all boundary samples are lower than the flat area, then the flat area represents a peak. If the boundary contains at least two higher contiguous regions that do not touch each other, then the flat area is a saddle. (Another way to imagine this is that in order to walk

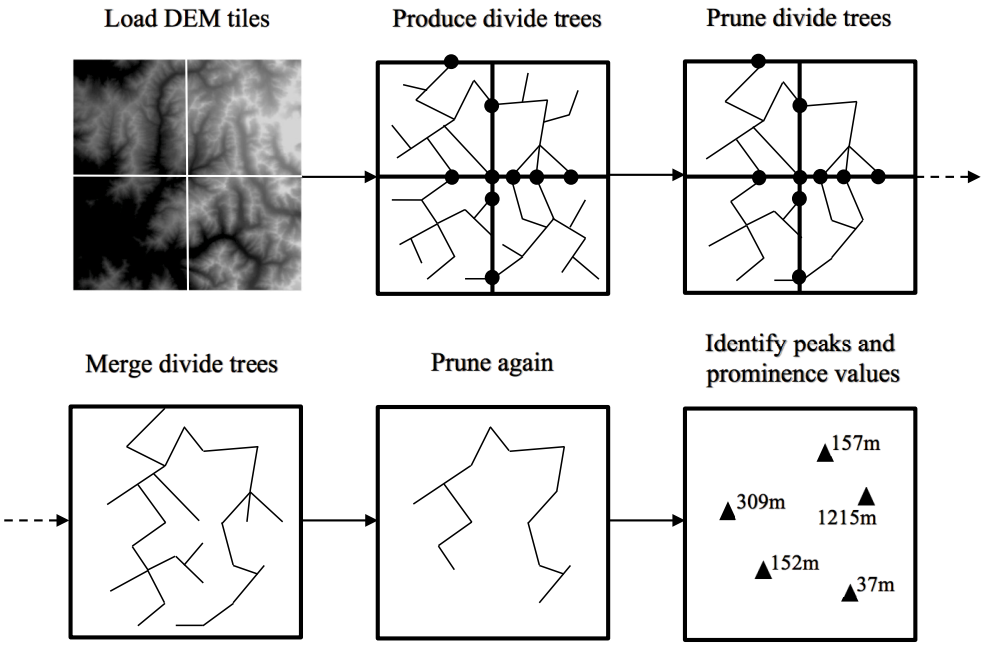

Figure 2: A schematic description of the prominence algorithm. Nodes of the trees are peaks and saddles in the DEM. Dark circles represent runoffs that will be used to merge neighboring tiles.

from one of the higher boundary regions to the other, one must pass through the saddle). It is also possible for the flat area to be neither a peak nor a saddle. In any case, we generate a unique identifier for the flat area and write it to the corresponding entries in the domain map. For each flat area, we store a structure with its type (peak, saddle, or other), a numerical identifier, elevation, and in the case of a saddle, the location of the highest points in its two boundary areas.

Some very large saddles have thousands of higher regions in their boundaries. To simplify what follows, we split up saddles with N higher boundary regions into $N-1$ logical saddles that each have exactly two higher regions. The particular selection of higher regions, and the precise location of each logical saddle within the original saddle's flat area are arbitrary, in that they do not affect the final prominence values. Earl (2016) went to great lengths to position these logical saddles so that their lines of steepest ascent would not cross, since he was focused on visual display. However, since presently we are looking only to enumerate prominent peaks, we create logical saddles in a straightforward way. Each logical saddle has as its two higher regions the highest boundary region from the original saddle, and one other higher boundary region. The location of the logical saddle is chosen to be the point in the flat area closest to the midpoint between these two boundary regions.

Next, for each saddle, we walk up the lines of steepest ascent until they each encounter a peak identifier in the domain map. Where this walk encounters other non-peak flat areas, we must again find the boundary of the flat area, identify the highest point in the boundary, and continue the uphill walk from there. There are three possible outcomes of this uphill walk:

1. Both ascent lines reach the same peak. This saddle cannot possibly be the key saddle of any peak, so we discard it.

2. The ascent lines reach different peaks, and there is currently no path in the divide tree between the two peaks. In this case we add two edges to the divide tree, between the saddle and each of the two peaks.

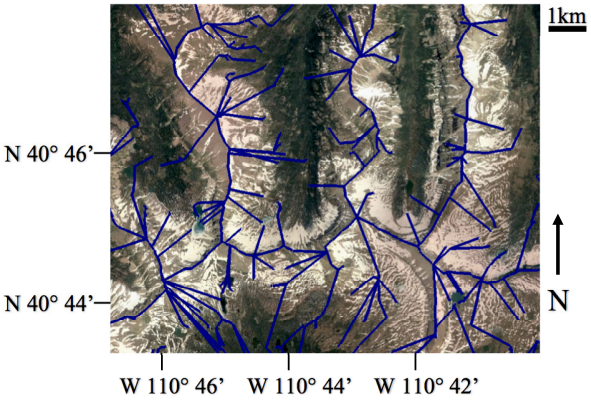

Figure 3: The divide tree in a mountainous region, showing how edges tend to follow ridge lines.

3. The ascent lines reach different peaks, but there is already a path in the divide tree between them. Adding new edges to the tree as above would create a cycle. We examine all of the saddles in the potential cycle and discard the lowest one, along with its two adjacent edges. Morphologically, this lowest saddle is a basin saddle, or the point where water would first spill out if the land inside the cycle were filled with water. Such a point cannot be a key saddle, as there is an alternative path connecting any two peaks that stays at a higher elevation, contradicting the definition of a key saddle.

Figure 3 illustrates the divide tree in a mountainous region.

2.3.2 Merging divide trees. To compute prominence over larger areas, we must merge the divide trees from adjacent tiles. The key to doing this is to recognize that by construction, the divide tree has only peaks at its leaves, since edges were always added between a saddle and the two peaks that were reached by uphill walks. Thus, to merge trees, we must identify potential saddles along their tiles' shared boundary.

When processing each tile, we scan its edges, looking for high points while considering only edge sam-

ples. Where the high point is in a flat area larger than a single sample, a representative point can be picked arbitrarily but deterministically, so that adjacent tiles will choose the exact same representatives. We refer to these high points as *runoffs*. A runoff is like half a saddle: the lower points along the tile edge separate the samples in neighboring tiles into two zones that don't touch each other. If both zones contain higher adjacent land, then the runoff is indeed a saddle, one that could not be detected by examining the samples in either neighboring tile by itself.

We also always generate a runoff at each of the tile's four corners. This is because it is possible for a saddle to exist at a corner regardless of the sample values in a single tile. Only after all four neighboring tiles are examined can we resolve the status of a runoff at a corner.

For each runoff, we store its location, elevation, and the peak reached by walking the line of steepest ascent. It is possible, and in fact common, for the sample location of a runoff to also contain a potential peak. For each runoff, we record whether its location corresponds to a peak's flat area in the domain map. The status of the runoff (a potential saddle), and that of the potential peak will be resolved when the tile's divide tree is merged with its neighbour.

Because we have guaranteed that the DEM is seamless, tiles that share an edge will find runoffs at exactly the same locations. Suppose that we have two tiles A and B that share an edge. First, we create a new divide tree that comprises the divide trees for A and B with no edges yet connecting them. Next, we find pairs of coincident runoffs and process each pair in turn. We replace the runoff pair with a new saddle and attempt to add it to the merged divide tree, exactly as we did when building the original divide trees. Sometimes the saddle will persist, either because its addition doesn't create a cycle, or because some other saddle in the cycle is lower, and the two trees will be spliced together at this point. If the runoff pair occurred at the corners of the tiles, then we need to preserve a runoff at this location in the merged tree until all four neighboring tiles have been merged in.

Recall that we recorded whether each runoff was within the flat area of a peak. Such peaks are provisional, in that we don't yet know whether there is higher ground bordering the flat area in a neighboring tile. When we examine a runoff pair, if both runoffs are marked as within the flat area of a peak, then the two flat areas are actually part of the same, legitimate peak, and we can combine the two peaks into one. If only one runoff of the pair is within a peak's flat area, then the flat area has higher bordering samples in the other tile, and the peak is removed.

2.3.3 Computing prominence. Proceeding from the definition, we can now compute the prominence of a peak by walking the divide tree, finding the highest saddle that is encountered along any path through the tree that ends at a higher peak, and then subtracting that saddle's elevation from the peak's elevation. In practice, this can be done more easily by first constructing a *prominence island tree* from the divide tree. In a prominence island tree, each peak is connected through its key saddle to its prominence island parent. A peak's prominence island is the area inside the lowest contour including no higher peak. A peak's prominence island parent is the highest point of the contour of the same elevation that meets its prominence island at its key saddle. Conceptually, it is on the opposite side of the saddle from the peak.

To compute the prominence island tree, we sort the divide tree by moving each peak upward until a higher parent is found. Suppose we have a section of the graph with three peaks $A \to B \to C$, where A is taller than B. We will move A above B, and we have a choice for B's new parent. If the AB saddle is lower than the BC saddle, we reparent B to A, otherwise we leave its parent as C. This ensures that the connection between B and its parent in the tree will contain its key saddle after the entire tree is sorted.

2.3.4 Pruning divide trees. A single square degree tile at ten meter resolution can contain over one million peaks, many of them only one meter high. It would not be feasible to merge many divide trees of this size. We must first *prune* the divide tree, removing all peaks whose prominence can be shown to be

below the threshold of interest P, which is 31m (100 feet) in this analysis. Since we need to preserve the peak-saddle-peak topology of the divide tree, we can only remove a peak if we can find a saddle connected by an edge to the peak, where the saddle cannot be the key saddle of a peak with prominence at least P. We continue removing such peak/saddle pairs until there is nothing available to remove.

To determine whether a peak can be removed, we compute its prominence as described earlier. If the peak's key saddle is not present in the divide tree, then it is not yet safe to remove.

To determine whether a saddle can be removed, we walk from each peak A along the divide tree to a higher peak. To the lowest saddle encountered on the walk, we assign a value equal to the difference between its elevation and A's elevation, which we can informally call the prominence of the saddle. If the walk reaches a runoff before a higher peak, then this saddle is assigned infinite prominence, because there could be a peak of unbounded elevation in another tile on the far side of the runoff. We cannot remove the saddle in this case, because it could be the key saddle of a peak we cannot yet see.

For each tile, we output the raw divide tree (in case it may be useful later), and the divide tree pruned to 31m. Merging occurs only on the pruned versions. The merging operation can leave behind peaks with prominence less than the threshold of interest, especially near the tile's edge, where the presence of a runoff prevented a small peak from being pruned earlier. Thus, it is useful to prune again after each merge operation. Figure 4 shows the effect of pruning a typical tile's divide tree.

2.3.5 Implementation. Several C++ programs were developed to calculate prominence. The first calculates the divide tree for a given range of latitudes and longitudes, and outputs both the original divide tree, and a pruned version. The second merges an arbitrary number of divide trees, prunes the result, and outputs a single, larger divide tree, and a text file with prominence values and key saddle locations. The first program can process tiles in parallel in a thread pool, while the second runs serially.

We divide the world into regions and process each independently in order to keep the runtime small. Tiles, however, cover one square degree and do not necessarily match our desired region boundaries. Thus, as a final step, a utility program takes as input a set of result peaks, and filters them to the interior of a polygon. This allows to produce the results for, say, Africa, without pulling in a small set of peaks from across the Strait of Gibraltar in Spain. These peaks could have incorrect prominence values unless we included all of Eurasia in our analysis of Africa, which we are trying to avoid.

III. Results

3.1 Isolation

A computer program written in C++ ran for approximately 10 hours on a modern laptop and produced a text file containing 24.7 million peaks with at least 1 kilometer of isolation. The record for each peak contained its location, the location of its ILP, its elevation, and its isolation in kilometers. After sorting the peaks by isolation and examining the top 1000 entries, several spurious entries were found that corresponded to single-sample errors in the underlying DEM. These so-called "spikes" were corrected in the DEM, and the tiles containing them were reprocessed.

The Peakbagger Web site contained several lists of peaks with isolation above certain threshold values. For most states in the Western U.S. there was a list with a threshold of 25 miles (40 km), and for Alaska, 100 miles (161 km). Among these lists, 3 peaks were found to have an ILP less than 25 miles away, and thus they were removed from the list. One peak in California (Mount Ritter) was found to have more than 25 miles of isolation, and one peak in Alaska was found with more than 100 miles of isolation. Using the results of this analysis, new lists were constructed for each continent with an appropriate isolation threshold, as well as a global list of 462 peaks with at least 300 kilometers of isolation. In addition, an ILP location and true isolation value ware added to all 60,000 entries in the Peakbagger database with

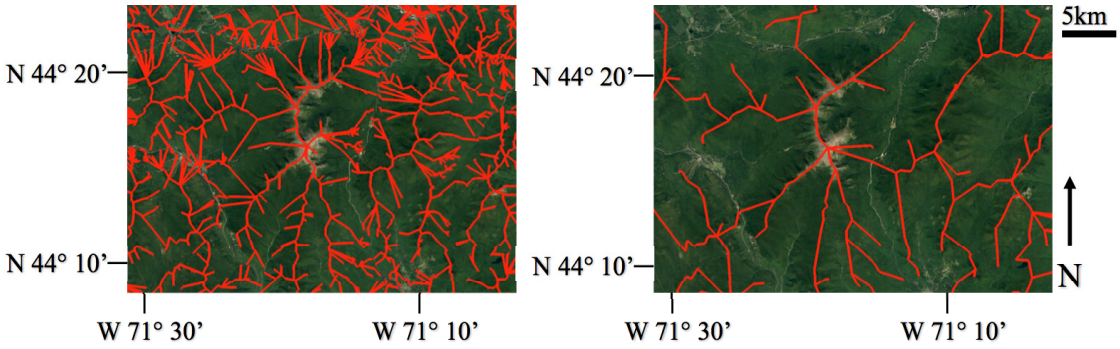

Figure 4: Original divide tree (a) and tree pruned to 31m of prominence (b) around Mount Washington, New Hampshire.

at least 1 kilometer of isolation.

This analysis focused on coverage, given that the isolation for most major mountains of the world was already known. However, it is instructive to compare the results with the previous most accurate list of the 40 most isolated peaks shown in Table 1.

There is a slight trend toward lower values in the new analysis, due to closer higher ground being found. Jarvis High Point illustrates one of the dangers of using surface models that include treetops. Its computed ILP is on Washington Island in Kiribati, which is a low, forested coral atoll. Although the highest ground elevation there is about 5 meters, the trees in the surface model bring it up over Jarvis's high point of 7 meters, "capturing" its high isolation value. With this exception, the entries in the table withstand scrutiny on topographic maps. A reordering is implied among the top 40 peaks, without any peaks dropping off the list.

3.2 Prominence

The world was divided into regions and each processed independently according to the scheme in Table 2.

Splitting the world this way kept the maximum size of the divide tree within reason, but it did cause erroneous prominence values for the highest point of each region cut on a land boundary. For example, Kilimanjaro's prominence parent is in Asia, which was analyzed separately from Africa. As a result, Kilimanjaro was incorrectly thought to be the high point of its land mass, and its prominence is too large by the height of its key saddle near the Suez Canal. Other mountains affected in the same way are Denali, Mount Whitney, and Pico de Orizaba. However, because the key saddles and prominence of these mountains are already well-known, we considered this to be only a minor defect.

Further difficulties arose where data sets of differing resolutions abutted. This occurred on the borders between the U.S. and Canada, U.S. and Mexico, and Mexico and Guatemala. When processing the lower resolution area (Canada, Mexico, and Guatemala respectively), it was possible to extend the range far enough into the higher resolution area to cover the key saddles of the major peaks in the lower resolution area, and thus the prominence values are correct. However, this was not possible when analyzing the higher resolution areas: some peaks have their key saddles outside the extent of the data. This caused overly high prominence values in the western U.S. near the Canadian border, and at the border of Mexico and Guatemala. The U.S.-Mexico border was much less affected because the high-resolution data extended far enough into Mexico to capture the important key saddles. It remains an open question how to combine DEMs of different resolutions with this algorithm.

Regions were processed in sequence over the course

Table 1: The 40 most isolated peaks in the world.

Rank	Peak name	Previous isolation (km)	New isolation (km)	New rank
1	Mount Everest	N/A	N/A	1
2	Aconcagua	16518	16518	2
3	Denali	7450	7450	3
4	Kilimanjaro	5510	5510	4
5	Puncak Jaya	5262	5235	5
6	Vinson Massif	4911	4928	6
7	Mont Orohena	4128	4128	7
8	Mauna Kea	3947	3947	8
9	Gunnbjørn Fjeld	3254	3255	9
10	Aoraki / Mount Cook	3139	3139	10
11	Thabana Ntlenyana	3004	3003	11
12	Maunga Terevaka	2832	2831	12
13	Mont Blanc	2812	2812	13
14	Piton des Neiges	2767	2766	14
15	Klyuchevskaya Sopka	2747	2750	15
16	Pico de Orizaba	2690	2690	16
17	Queen Mary's Peak	2662	2662	17
18	Mount Whitney	2649	2649	18
19	Gunung Kinabalu	2513	2509	19
20	Mount Elbrus	2470	2469	20
21	Pico de Banderia	2344	2342	21
22	Mont Cameroun	2338	2335	22
23	Mount Paget	2267	2205	25
24	Mauga Silisili	2245	2245	23
25	Nevado Huasacarán	2196	2207	24
26	Anai Mudi	2109	2090	26
27	Jebel Toukbal	2078	2078	27
28	Mount Fuji	2077	2077	28
29	Mawson Peak	2046	2026	29
30	Emi Koussi	1999	1999	30
31	Mount Mitchell	1914	1914	31
32	Gunung Kerinci	1904	1904	32
33	Agrihan High Point	1898	1898	33
34	Mount Kosciuszko	1894	1894	34
35	Olavtoppen	1851	1851	35
36	Jarvis High Point	1849	559	See text
37	Mascarin Peak	1827	1805	39
38	Green Mountain	1813	1798	40
39	Goro Narodnaya	1835	1835	37
40	Yushan	1815	1815	38

Table 2: Regions used in the prominence calculation, with the terrain resolution in each.

Region	Description	DEM resolution (m)
Africa	Cut at Suez Canal; Madagascar	90
Antarctica	Continent and South Pacific islands	90
Australia	Mainland and Tasmania	90
Central America	Belize to Panama Canal	90
Eurasia	Cut at Suez Canal	90
Greenland		90
Islands	Pacific Islands, Indonesia, Japan, New Zealand	90
Mexico		30
South America	Cut at Panama Canal	90
Alaska	Includes western Yukon and northwest British Columbia	90
Canada		30
USA	Lower 48 states and Hawaii	10

of several weeks on a modern laptop computer. Typically, a region would be run, then the pruned divide trees for each 10-degree band of latitude would be merged, and then the results from all of the merged bands would be merged again into one divide tree for the entire region. The largest tree was that for Eurasia, which covered 7689 square degrees. The representation of its divide tree in text format totalled about 300 megabytes.

Table 3 summarizes the number of peaks found at each conventional prominence threshold.

Because the source data for Antarctica had only 200 meter spacing, we did not attempt to calculate peaks with 31 meters of prominence there. The asterisked total for the U.S. includes three peaks near the Canadian border whose prominence is actually much less, due to their prominence parents being outside

Table 3: Count of peaks found above each traditional prominence threshold, by region.

Region	31m (100')	91m (300')	610m (2000')	1500m
Africa	1,107,905	181,592	1317	87
Antarctica	N/A	4672	283	40
Australia	148,831	19,998	113	2
Central America	83,275	14,438	265	26
Eurasia	3,443,244	811,816	14,189	604
Greenland	34,927	18,983	1047	41
Islands	393,938	85,195	1978	189
Mexico	270,447	52,070	757	26
South America	1,068,390	191,154	3688	205
Alaska	134,390	50,418	1744	75
Canada	588,277	125,026	4274	115
U.S. (outside Alaska)	520,413	78,664	1240	67*
Total	**7,798,709**	**1,634,026**	**30,895**	**1477**

the high-resolution DEM.

Where the source data was a surface model, meaning most places outside of the U.S., Canada, and Mexico, results at the level of 31 meters of prominence are likely quite noisy, incorporating forest canopies and even tall buildings. The purpose of going to this low threshold was to demonstrate the power of the method. The 91-meter prominence level is a much more realistic minimum in these areas, and even then, values near the minimum threshold would need to be checked against topographic maps to validate them.

As with the isolation calculation, the goal of this analysis was to extend coverage over the globe, but again it is interesting to compare with the previously known most prominent peaks in the world. A total of 1524 ultras (with 1500 meters of prominence) were previously known, and 13 new ones were found, as shown in Table 4.

The reason these ultras had not be found previously is either that the voids they inhabited had not yet been filled at the time of earlier analysis, or their key saddles were too far away for the previous analyses to find. In addition, the current analysis did not identify 50 previously known ultras, all with prominence just over 1500m. Most likely this is due to the tendency of DEMs to underestimate summit elevations, as mentioned earlier.

Source code, text files with full results, and links to online visualizations are available online at github.com/akirmse/mountains.

IV. Conclusion

Efficient algorithms are available to compute isolation and prominence values down to the resolution that is supported by the underlying terrain data. Void filling is critical to analysis of mountainous regions, and we have described the first DEM with sufficient resolution and coverage to perform a global calculation of isolation down to 1 kilometer, and prominence down to 30 meters.

An obvious future task is running these algorithms over a higher resolution DEM with global coverage. The best immediate candidate is the WorldDEM terrain model from the TanDEM-X mission (Airbus, 2015), which has 12m sample spacing and 10m vertical accuracy. This data set is still being built, and is not freely available, so we were not able to evaluate its coverage and quality. Processing this volume of terrain data would require substantial computational parallelism.

To further improve the accuracy of the results in the future, we must find a way to incorporate known peak and saddle elevations, many of which exist only

Table 4: Newly discovered ultra-prominent peaks

Name if known, country	Location (degrees)	Elevation (m)	Prominence (m)
Borgtinden, Greenland	68.8517, -28.2375	3297	1790
Uludaz, Turkey	37.45, 36.655	2233	1720
Greenland	66.2017, -37.2217	1940	1609
Molkooh, Iran	35.5208, 58.8342	2958	1560
Niujiao'an, China	36.8175, 111.9717	2555	1539
Terektinsky Ridge, Russia	50.2442, 86.4892	2911	1534
Cerro Carreras, Argentina	-41.5042, -71.2358	2364	1531
Xiaowutai Shan, China	39.9417, 115.0433	2836	1519
Chile	-46.5692, -73.0783	2300	1512
Chile	-54.565, -70.4292	1860	1511
Wuling Shan, China	40.5983, 117.4808	2099	1508
Magefjeld, Greenland	71.9775, -53.6833	1777	1501

in survey data that is not in digital form. One stopgap possibility is to use recent advances in computer vision to import the location and elevation of points from scanned and geolocated paper maps in bulk. Another is to build seamless, high-resolution DEMs from aerial LIDAR surveys. Either method would allow us to better differentiate nearby peaks with similar summit elevations.

References

Airbus Defence and Space (2005). *WorldDEM technical product specification 2.0.* Geo-Intelligence Program Line.

Čomić L, De Floriana L and Papaleo L (2005) Morse-Smale decomposition for modeling terrain knowledge. In: Cohn AG and Mark DM (eds) *Spatial Information Theory. COSIT 2005. Lecture Notes in Computer Science, Vol 3693.* Berlin, Heidelberg: Springer.

de Ferranti J (2011) ASTER Digital Elevation Data. Available at http://viewfinderpanoramas.org/reviews.html#aster (accessed June 10, 2017).

de Ferranti J (2016) Digital Elevation Data. Available at http://viewfinderpanoramas.org/dem3.html (accessed June 10, 2017).

Deng Y and Wilson J P (2008) Multi-scale and multi-criteria mapping of mountain peaks as fuzzy entities. *International Journal of Geographical Information Science*, 22(2), 205-218.

DiMarzio J P (2007) *GLAS/ICESat 1 km Laser Altimetry Digital Elevation Model of Greenland, Version 1.* Boulder, Colorado: National Snow and Ice Data Center. Available at http://nsidc.org/data/dems (accessed June 10, 2017).

Earl E (2016) WinProm. Available at http://github.com/edwardearl/winprom (accessed June 10, 2017).

Fisher P, Wood J, and Cheng T (2004) Where is Helvellyn? Fuzziness of multi-scale landscape morphometry. *Transactions of the Institute of British Geographers*, 29, 106-128.

Fry S (1987) Defining and sizing-up mountains. *Summit*, Jan-Feb, 1987, pp. 16-21.

Gesch D, Oimoen M, Greenlee S, Nelson C, Steuck, M, and Tyler D (2002) The National Elevation Dataset. *Photogrammetric Engineering and Remote Sensing*, 68, 5-11 (accessed June 10, 2017).

Gesch DB, Oimoen MJ, and Evans GA (2014), Accuracy assessment of the U.S. Geological Survey National Elevation Dataset, and comparison with

other large-area elevation datasets—SRTM and ASTER: U.S. Geological Survey Open-File Report 2014–1008, 10 pp. Available at http://dx.doi.org/10.3133/ofr20141008 (accessed June 10, 2017).

Helman A (2005) *The Finest Peaks - Prominence and Other Mountain Measures*. Victoria, Canada: Trafford Publishing.

Jurgalski E (2009) Short history of orometrical prominence. Available at http://www.8000ers.com/cms/en/prominence-mainmenu-179.html (accessed June 10, 2017).

Kirk J (2016) ListsOfJohn. Available at http://listsofjohn.com (accessed June 10, 2017).

Liu H, Jezek KC, Li B, and Zhao Z (2001) *Radarsat Antarctic Mapping Project Digital Elevation Model, Version 2*. Boulder, Colorado: NASA National Snow and Ice Data Center Distributed Active Archive Center.

Llobera M (2001) Building past landscape perception with GIS: Understanding topographic prominence. *Journal of Archaeological Science* 28(9): 1005-1014.

Maizlish A and Slayden G (2004) Isolation. Available at http://www.peaklist.org/theory/orometry/article/Orometry_9.html (accessed June 10, 2017).

Martin A (1994) *County High Points*. Tucson, Arizona: Old Adit Press.

Meeus J (1991) *Astronomical Algorithms*. Richmond, Virginia: Willmann-Bell.

Munro H (1891) Tables giving all the Scottish mountains exceeding 3,000 feet in height. *Scottish Mountaineering Club Journal* 1(6), 279-281.

Podobnikar T (2012) Detecting mountain peaks and delineating their shapes using digital elevation models, remote sensing and geographic information systems using autometric methodological procedures. *Remote Sensing* 2012, 4, 784-809.

Rodriguez E, Morris CS, Belz J, Chapin E, Martin J, Daffer W, and Hensley S (2005) *An Assessment of the SRTM topographic products, Technical Report JPL D-31639*. Pasadena, California: Jet Propulsion Laboratory.

Slayden G (2016), World peaks with 1000 km of isolation. Available at http://peakbagger.com/list.aspx?lid=301 (accessed December 25, 2016).

Tomko M and Purves R (2010) Matterhorn on the horizon: Identification of salient mountains for image annotation. In: *proceedings of the GIS Research UK 18th Annual Conference GISRUK 2010*, University College London (UCL), 14–16 April 2010, 359–362.

US Geological Survey (2004), *Shuttle Radar Topography Mission, 3 Arc Second scenes*, College Park, Maryland: Global Land Cover Facility, University of Maryland.