Choice Set Generation for Modeling Scenic Route Choice Behavior Using Geographic Information Systems

Word Count: 5497 plus 6 figures/tables at 250 words each = 6997 words, number of references: 35

Revision date: March 6, 2015

Majid Alivand
University of Florida, Geomatics Program
3205 College Avenue
Ft. Lauderdale, FL-33314
Phone: (954) 577-6392
Fax: (954) 475-4125
E-mail: malivand@ufl.edu

Hartwig H. Hochmair
University of Florida, Geomatics Program
3205 College Avenue
Ft. Lauderdale, FL-33314
Phone: (954) 577-6317
Fax: (954) 475-4125
E-mail: hhhochmair@ufl.edu
ABSTRACT
Finding a scenic route between two locations is a common trip planning task, in particular for tourists and recreational travelers. For the automated computation of a scenic route in a trip planning system it is necessary to understand which attributes of a route and its surroundings are associated with attractive scenery. One way to answer this question is to use a route choice model framework which requires a choice set of realistic paths for traveling a scenic route. This study proposes a constrained enumeration type algorithm for computing scenic alternatives using a branch-and-bound approach and a labeling algorithm in combination with link elimination to generate a scenic route choice set. These algorithms use Volunteered Geographic Information (VGI), more specifically shared geotagged photos from Panoramio, to generate path alternatives. The results show that the proposed branch-and-bound algorithm replicates observed travel behavior better than the traditional deterministic minimum-cost based algorithms in terms of overlap and consistency. Also, for the proposed branch-and-bound method the average number of scenic clusters is almost identical to that found along observed scenic routes, which outperforms all other evaluated algorithms in this aspect.
INTRODUCTION
Online trip planners have become part of our daily lives. They allow users to find the optimal path between two locations based on a specific criterion such as travel time, travel distance, or number of turns between the predefined locations in street networks (28). A research area which is currently gaining interest among the transportation community is the computation of optimized routes for tourists and recreational travelers (15, 18, 32). A shortest and in general single criterion optimal path is not what a tourist typically needs when planning a trip. That is, the computed path should not only be short or fast but at the same time satisfy other conflicting optimization criteria, such as passing through landscapes (30). Thus, the computation of a scenic route is inherently a multi-criteria problem.

An important question which arises in this context is what constitutes a scenic route. In other words, what types of criteria are involved in a traveler’s process of determining the scenic value of a route? One way to answer this question is by analyzing how tourists and recreational travelers choose scenic routes based on reported trips, and relate their choices to explanatory variables. This can be approached through route choice models. The results of these models indicate the significant attributes with their coefficients, which can be used for the computation of scenic routes in automated route planning systems. Route choice models require a heterogeneous choice set of realistic routes. Scenic routes exhibit unique characteristics, for example, side-trips branching off the main-route direction to a viewpoint located on a dead-end, which cannot be replicated through typical minimum-cost based path generation algorithms. This paper introduces three path generation approaches to build choice sets for identifying scenic route attributes. The algorithms use clusters of geotagged images from Panoramio, a shared photo service, to build route alternatives. Shared photos are an example of Volunteered Geographic Information (VGI) (14), which complements traditional geo-data sources, such as aerial photographs or census data, through voluntary data contributions on Web 2.0 applications. The proposed path generation algorithms are implemented in a Geographical Information System (GIS) framework. Generated choice sets will be compared to those created through traditional choice set generation methods, such as labeling, link elimination and link penalty, with respect to coverage, consistency and number of scenic point clusters along the route.

RELATED CONCEPTS
Previous studies on scenic routes can be divided into those highlighting the importance of certain attributes for perceiving a route as scenic, and those that introduce algorithms for computing scenic routes. The significance of scenic attributes in route selection criteria using route choice models was examined in (6, 31). These studies conclude that the scenic attribute contributed significantly to the utility function of the route choice model, even for home to work commutes (31). The scenic beauty of routes using visibility analysis and evaluation of landscape photos was estimated in (7, 32, 33). Algorithms for the computation of scenic routes include a single objective shortest path methodology on a network with modified edge weights (18), or the use of Internal Path Discovering (IPD) and Bellman-Ford algorithms with VGI data sources (21, 33). A recent study that used a binary choice model found that water bodies, mountains, and forests are contributing to route scenery, and that the density of Panoramio images along a route is positively associated with chosen scenic routes (2). In that particular study, each choice set consisted of two alternatives, i.e., the chosen scenic and the corresponding fastest route.
Choice Set Generation
The exhaustive approach for path generation unrealistically assumes that all physical routes connecting origin and destination of a trip are considered by a traveler. Thus, selective approaches have been proposed that account for deterministic and probabilistic procedures to generate the paths. The generated choice set should exclude routes that travelers would not consider in their route selection process, as well as highly similar paths which travelers do not differentiate. Instead, it should include relevant and heterogeneous routes that different travelers choose (25). Prato (26) categorizes path generation algorithms into four categories, which are deterministic shortest path-based methods (e.g. link elimination), stochastic shortest path-based methods (e.g. simulation approach), constrained enumeration methods (e.g. branch-and-bound algorithms), and probabilistic methods (e.g. the implicit Availability/Perception (IAP) model).

It has been shown that choice set size and composition affect model estimates and choice probabilities for several model specifications when Multi-Nomial Logit (MNL) models or its modifications are estimated (5, 25, 8). However these studies also show that the inclusion of unattractive routes in the choice set neither distorts the demand predictions nor seriously influences the computational efficiency (10), so that the inclusion of some unattractive routes is acceptable.

ANALYZING SCENIC ROUTE CHOICE BEHAVIOR
Selecting a scenic route among several route alternatives involves multi-attribute decision making (MADM). This means that a traveler chooses between a limited number of decision alternatives based on their attributes. Statistical elicitation of criteria that are considered by a traveler in the selection of scenic routes typically employs route choice models, using a realistic route choice set. We assume that a choice set generation method that goes beyond traditional choice set generation approaches will allow to extract additional, previously unidentified selection criteria for scenic routes, such as the role of dead ends or loops. One of the algorithms introduced in this study is a constrained enumeration approach that generates a choice set consisting of several scenery related routes alternatives. The nature of these routes reflects characteristics of observed chosen scenic routes, such as passing through spatial clusters of shared photos. Besides this approach, the study introduces and evaluates two other algorithms that use shared photos for route generation.

VGI Data Sources and Study Area
VGI has been facilitated through the recent development of Web 2.0 applications and the integration of GPS sensors into mobile devices such as smartphones and digital cameras. It is a freely available Web-based data source that can be used to elicit people’s travel behavior. The state of California has been chosen as the study area for testing the various path generation methods since it provides a rich set of VGI data sources, mainly through numerous tourists traveling to California each year.

Two types of VGI data were utilized for the extraction of section routes. Geo-tagged photos from the photo sharing Websites Panoramio and Flickr are the first type of data sources. Flickr and Panoramio are Web 2.0 portals which host digital photos with spatial information and timestamps. Most of this information can be accessed through designated API’s. On these Websites, contributors upload photos primarily from places which they consider interesting or scenic, which can be used to extract traveled scenic routes (21, 34). The second type of VGI data used in this study is travel-diary Websites, such as RouteYou, EveryTrail, and MyScenicDrive,
where users upload traveled routes (e.g. GPS tracks). Their major advantage is that they provide the complete route geometry together with trip origin and destination. A total of 96 scenic routes were extracted from these two VGI data sources, which represent the set of chosen scenic routes in this paper. Their length ranges between 10 km and 800 km. Details about the extraction of scenic routes from photo-sharing and travel diary Websites for the California test area are described in (I).

**Observed Travel Behavior**
A route generation algorithm should create route alternatives that replicate the chosen scenic routes as far as possible but also show some diversity in characteristics while being comparable to the chosen scenic route. A detailed analysis of chosen routes with respect to different spatial, temporal and directional constraints is required in order to replicate the chosen scenic routes as closely as possible.

Analyzing the 96 chosen routes reveals that the travel time of the chosen scenic routes is on average 90% longer than corresponding fastest route. It also shows that travelers willingly take detours and dead-end routes to visit a scenic spot or landmark. The characteristics of observed scenic routes, which are a result of the traveler’s route choice behavior, are different from those of shortest paths. Even with no map or GPS available but only a general direction to the destination given, travelers typically show at decision points preference for road segments that lead them initially as fast as possible to the destination. This includes roads segments which are most in-line with the direction to the destination (I7), that conserve linearity throughout the trip towards the destination (I1) and that are part of a long and continuous road rather than a shorter and winding road (4). However, a significant portion of the observed 96 scenic routes reveals different route characteristics around the trip origin. That is, the direction of chosen road segments deviates sometimes heavily from the direction towards the destination (i.e. pointing away from the destination), and more curved roads are chosen than what would be the case for fastest routes. Especially the proposed cluster based branch-and-bound route generation algorithm provides enough flexibility to replicate these specific characteristics of observed scenic routes.

**GENERATION OF SCENIC ROUTE ALTERNATIVES**
VGI provides insight into the perceived attractiveness of locations in an environment. Three algorithms are proposed for the generation of scenic alternatives that make use of the location of geotagged photos.

The proposed algorithms are based on images posted on photo sharing Websites. In Panoramio all images are geotagged, whereas in Flickr only about 3-4% are. Further Panoramio images provide a better positional accuracy than Flickr images (35), and all Panoramio images are taken outdoors, which is not the case for Flickr images. Because of these reasons, we chose Panoramio images as a basis for generating alternative scenic routes. Around 250000 photo footprints were downloaded from the Panoramio Website for further analysis.

Geotagged photos can be used to identify scenic hot spots through clustering. These clusters can then be used as candidate waypoints in scenic route generation algorithms. The TomTom 2013 dataset for California was used for network analyses, and computed travel times based on posted speed limits were used to compute travel times. All evaluated path generation algorithms were coded using ArcObjects in the .NET framework and executed in combination with the ArcGIS Network Analyst extension.
Identification of Scenic Points
Scenic points were extracted from the location of geotagged Panoramio images through three processing steps, which are image filtering, clustering, and snapping cluster centroids to the road network.

Image filtering
Since in this study, shared photos are used for routing purposes, only Panoramio images within a 50 m buffer around the road network were used. Although all Panoramio images show outdoor scenes, not all of them resemble what one would typically associate with attractive locations. That is, photos are sometimes taken at less scenic locations such as restaurants, gas stations, bus stops, or highways. Various methodologies have been proposed to automatically remove such images from the collection of selected photos (16, 29). We applied the principle introduced in (16) which states that a shared outdoor photo indicates a scenic location only if another user posts a photo nearby, i.e. within a given threshold distance. That threshold distance should be set smaller in urban areas than in rural areas, taking into account that in rural areas the scenic locations are typically visible from larger distances than in urban areas. Using this filter, 84445 Panoramio photos were retained for further analysis.

Clustering
The nearest neighbor hierarchical clustering method was applied on the filtered set of Panoramio images in the CrimeStat software package (20) to identify hot spots. The radius and the minimum number of points per cluster are the two parameters which affect the size and the number of clusters. We determined parameter settings in an exploratory manner that reasonably balanced cluster size and the number of clusters. This resulted in 4134 photo clusters.

Cluster snapping
For routing purposes each cluster was replaced by its centroid point, which was then snapped to the road segment with the highest road category available nearby the original centroid location. Highways and ramps were excluded in the snapping process.

Proposed Algorithms for Generating Scenic Alternatives
Using scenic cluster points identified in the previous steps, three methods are proposed to generate scenic alternative routes. The first method falls into the category of constrained enumeration methods, which relies on the behavioral, cognitive and perceptual assumptions that travelers choose routes according to behavioral rules other than a minimum cost path (26). The approach aims to create a path that passes through scenic cluster points and satisfies various spatio-temporal constraints to be followed in a step-wise building approach. The second method uses a labeling approach with reduced link cost in the vicinity of scenic cluster points, leading the least-cost algorithm to gravitate towards the scenic points along the path. The third method combines the second method with a link-elimination approach.

Branch-and-bound algorithm
The first method uses a branch-and-bound algorithm which generates paths that run through scenic points under consideration of spatio-temporal and directional constraints. Spatio-temporal constraints help to circumscribe individual activity-travel behavior and can be used to define potential path areas (PPA). Several GIS-based approaches have been proposed that incorporate such constraints in the definition of PPAs (22, 19, 27). These two-dimensional representations
for three-dimensional space-time prisms define realistic areas that can be reached by individuals under given spatio-temporal constraints, such as a limited travel time budget (TTB). We applied the overlay approach (27) to determine the PPA that limits the spatial extent of scenic route alternatives. The overlay approach divides the TTB via a constant time increment (we chose 2 minutes) into several combinations of travel times from the origin i to the destination j. All network links reachable from i and j are identified, followed by an intersection of the two sets to form a new set containing all common links. This two-step procedure is repeated for all remaining travel-time combinations. The PPA is then constructed from the union of all intersected sets of network links. We set the TTB equal to the travel time of the chosen scenic route plus 20 percent. Figure illustrates the steps for generating a PPA assuming a TTB of 20 minutes, based on a travel time of approximately 17 minutes along the chosen scenic route. For illustration purposes only three travel-time combinations are shown. Figure a shows for the three travel-time combinations the areas that can be reached from the origin and the destination given the time constraints. Resulting service areas are then intersected (Figure b) and appended (Figure c) in the GIS to generate the PPA for the origin-destination pair.

The path generation algorithm (Figure ) starts with the trip origin and selects in an iterative process new waypoints from available scenic cluster points which are added to the waypoint list until the destination is reached. To increase the variation of generated paths, the number of selected waypoints is randomly chosen between 2 and 6 before each path generation process.

Each time a new waypoint is searched, a spatial constraint is applied. This constraint identifies candidate scenic points within a distance range from the previous waypoint (or the origin) from which the next waypoint will be selected. For this task a multiple ring buffer is created (Figure , box a). The mean radius for that multiple ring buffer is computed through dividing the direct origin-destination distance by the number of waypoints plus one. The ring buffer is constructed by using an inner and outer radius of the mean radius times (1±0.2). All scenic points falling between the two radii are selected as candidate waypoints (box b). If no scenic points can be identified, the width of the ring buffer is stepwise increased until at least one candidate point can be identified (box c).

In order to guide the route from the origin towards the destination an additional directional constraint is applied in each step. This reflects a greedy behavior, i.e. a heuristics that moves the traveler in the general direction of the destination without consideration of future consequences for the remaining route (9, 17). Thus, in each step the azimuth from the latest waypoint to the destination and each scenic point in the ring buffer is computed (box d). Only scenic points which deviate less than a deviation threshold value from the direction towards the destination are retained as waypoint candidates (Figure b). Observed chosen routes were found to sometimes deviate heavily from the direction to the destination at the beginning of the route. To reflect such route preference, the directional constraints are less stringent at earlier waypoints near the trip origin. For example, at the trip origin candidate points in the ring buffer can have any deviation from the direction to the target, whereas with decreasing distance from the destination, the angular threshold becomes narrower. This means that the next waypoint is chosen from candidate points in the ring buffer that are more closely aligned with the destination direction (see circular arcs shown in Figure b). Further, only candidate points that are located in the same general direction as the direction followed from the previous waypoint are included, which avoids a bouncing back and forth between two neighboring regions with multiple scenic point clusters. One of the candidate points that satisfies these directional constraints is finally
chosen as the next waypoint and added to the waypoint list (box e). If the newly added waypoint is not close enough to the destination to complete the route, a new waypoint is searched (box f), otherwise the destination point is added to the list of waypoints, and the search for waypoints is complete (box g). The route alternative is then computed by connecting all points in the waypoint list through fastest route computation (box h).

For each origin-destination pair a maximum of 50 alternatives was created in this way. Figure illustrates for a chosen scenic route in San Francisco the iteration steps for generating a scenic route alternative, based on the algorithm depicted in Figure. In this example, three intermediate waypoints were identified. Figure b shows the candidate waypoints falling inside the ring buffer before the directional filter is applied.

Reduced link cost around scenic cluster points
This algorithm falls into the category of labeling approaches. The method first reduces the traverse cost of network edges in the vicinity of scenic spots, i.e. within a service area computed in the GIS. Next it computes the least-cost path (using travel time) between the trip origin and destination. A variation of this method has been used for trip planning applications, e.g. to compute routes along water bodies and parks (18), or routes with the highest Level of Service (LOS) for pedestrians (23). Different time thresholds were used for generating service areas around scenic spots in urban and rural areas, respectively. The travel time for network links that intersect with these service areas was reduced by 80 percent, which leads to routes that gravitate more towards scenic cluster points than with smaller cost reductions, such as 20 or 50 percent, which was also tested. This algorithm results in one alternative path per origin-destination pair.

Reduced link cost and link elimination
We combined this approach with a link elimination method, where a randomly chosen link from a major road along the computed fastest route is eliminated at each step, followed by computing a new alternative route, again based on the fastest route algorithm. A maximum of 50 alternatives were computed using this combination of reduced link cost and link elimination methods.

Route Evaluation and Removal
The proposed three methods together created a maximum of 131 different scenic route alternatives for a chosen scenic route. Since the reduced link cost method and its combination with link elimination are based on Dijkstra’s shortest path algorithm and link-elimination techniques present asymptotical behavior with respect to the ability to produce unique routes (i.e. they are limited in the number of different routes they produce) (24), the generated routes are reasonable with regard to the detour range and shape. However, the generated alternatives using the branch-and-bound algorithm through scenic cluster points sometimes contain undesirable loops and detours, which necessitate a subsequent route removal process. Figure a shows an example of a route generated with the branch-and-bound algorithm through scenic cluster points with an undesirable detour.

Such a route can be removed from the choice set through a time constraint. All routes that had a travel time up to the scenic route travel time plus 70% of the difference in travel time between scenic and corresponding fastest route were retained, which was derived from characteristics of observed VGI routes. As a result, around 40 percent of the generated scenic routes using the branch-and-bound algorithm were removed from the choice sets, including the generated route alternative shown in Figure a.
In some cases, generated routes fall within the acceptable detour threshold, but have undesirable loops. In order to remove these routes, first all the generated alternatives were converted to polygons to identify loops. Different parameters were used to determine if a loop was acceptable as part of a route. These parameters help to distinguish between loops caused by road network design along a realistic route, such as looped highway ramps, or loops caused by passing through a scenic spot in an undesirable way.

The first parameter used to classify loops is the shape factor (SF), which is computed as

\[
\text{shapefactor} = \frac{P_p^2}{A_p}
\]

where
- \(P_p\) is the perimeter (km) of the polygon enclosed by the street network
- \(A_p\) is the area (km\(^2\)) of the polygon enclosed by the street network

A higher SF value indicates a more elongated polygon. Figure b and c show the shape factors for the extracted loops from two generated routes. Loops caused by ramps and intersection areas (see loop 3 in Figure c) have a small shape factor compared to potentially undesired loops caused by scenic spots and backtracking along the same road (see Figure b and loops 1 and 2 in Figure c). Routes whose polygons have only small shape factors were therefore kept in the choice set, whereas those routes with larger polygon shape factors were further evaluated through additional parameters as follows.

If a scenic point is far from the start or endpoint of a loop relative to the direct distance between start and end point of the loop, this indicates that the route runs off the main direction towards the scenic spot, which is a common characteristic of scenic routes. Thus the direct distance between start and end point of the loop (\(d_1\)) and the direct distance between the scenic spot and either the start or end point of the loop, whichever is smaller, (\(d_2\)) was determined. The values were compared to loop characteristics of chosen scenic routes to keep those generated routes that replicate characteristics of chosen routes by travelers as closely as possible. In general an increased \(d_1\) also required \(d_2\) to increase to result in an acceptable route.

Figure b and c show two examples of routes with loops that were examined using these two parameters. For Figure b the scenic cluster point is near the main road and therefore most likely visible from the other side of the road without the need to reverse the travel direction and run a loop. This route was removed based on established distance thresholds. In Figure c the selection of the scenic spot creates three nested loops. That is, between the start and end point of loop 1 two more loops occur (loop 2 and loop 3). In such a situation, only the characteristics of the outer loop (loop 1) are considered for the loop evaluation, which in this case lead to retaining this route for further analysis.

OTHER METHODS OF ALTERNATIVE PATH GENERATION
We compared the generated scenic alternatives with results from other deterministic choice set generation methods that were applied to the same 96 origin-destination pairs.

The first applied method is the labeling approach which finds a least-cost path based on different individual route attributes. Based on route preferences determined in previous studies, we chose to minimize travel time (free-flow), distance, number of turns, travel time along highways, and number of road type changes along the route (considering highways, major road,
and local road as road classes). The labeling approach creates one route alternative per origin-destination pair and attribute.

Further, a modified version of link elimination method was applied. It searches iteratively for the next optimal path by removing one or more links from the computed least-cost path. We applied 10 iterations of the following two steps: (1) computation of the fastest route between origin and destination, (2) elimination of a random link belonging to the highest road category (highway if available, otherwise major road) from the current fastest route.

Next a link penalty approach was applied. It iteratively computes least-cost paths by imposing a penalty to links instead of removing them. We applied 15 iterations of the following two steps: (1) computation of the fastest route between origin and destination, (2) adding a time penalty of 20 percent to the links of the current fastest route. The link penalty and link elimination techniques present asymptotical behavior with respect to the ability to produce unique routes and consequently additional iterations give reduced gain in coverage.

Next, identical routes were removed from within generated choice sets.

EVALUATION OF CHOICE SET GENERATION ALGORITHMS

It is important for modeling purposes that choice sets are consistent with the observed travel behavior. One way to assess this characteristic is to check whether a choice set generated by different algorithms contains the actual chosen route. More specifically, we evaluate the generated choice sets based on coverage and consistency. Coverage measures the percentage of observations for which a path generation technique reproduces the actual behavior according to a certain overlap threshold $\delta$:

$$ O_{nr} = \frac{L_{nr}}{L_n} $$

where $L_{nr}$ is the overlapping length between the generated alternative by algorithm $r$ and the chosen route $n$, and $L_n$ is the length of the chosen route $n$. The coverage is the percentage of observations for which an algorithm generates a route that satisfies a threshold $\delta$ for the overlap measure:

$$ \sum_{n=1}^{N} I(O_{nr} \geq \delta) $$

where $I(\cdot)$ is the coverage function and is equal to 1 if its argument is true and zero otherwise.

The ideal algorithm will perfectly reproduce the chosen routes by replicating link by link all observed scenic routes, which would result in 100% coverage for a 100% overlap threshold. However, the actual algorithms only partially replicate the observed behavior, thus an index measure is used to measure the behavioral consistency of route generation methods with respect to the ideal algorithm by accounting for total overlap over all the observations:

$$ CI_r = \sum_{n=1}^{N} O_{nr,max} $$

where $CI_r$ is the consistency index of algorithm $r$, $O_{nr,max}$ is the maximum overlap measure obtained with the generated routes by algorithm $r$ for the chosen route $n$, and $O_{max}$ is the 100% overlap over all the N observations for the ideal algorithm.
Table provides for each implemented path generation algorithm the number of alternatives computed per origin-destination pair, as well as the number of unique routes generated through each path generation algorithm. It further presents coverage results with regard to overlap thresholds ranging between 90% and 60%. The individual labeling approaches perform worst and do not replicate the observed scenic route well, even for a 60% overlap threshold. The labeling approach with reduced link cost around scenic spots as one of the proposed algorithms also performs poorly. The proposed branch-and-bound algorithm provides the best coverage results among all the methods, followed by the link penalty approach and the proposed combination of reduced link cost and link elimination. The merged choice sets from all algorithms (bottom row) show the overall coverage results.

The right-most column in Table shows that the proposed branch-and-bound has the highest consistency of path generation methods, followed by the link penalty approach and the proposed labeling approach with reduced link cost combined with link elimination. The merged choice set performs best.

Besides overlap and consistency measures additional quality measures should be evaluated for generated routes that relate to route attractiveness. Thus, if available one could compute descriptive statistics for various potential scenic attributes, such as percentage of forest or coast line, for generated and observed scenic routes and compare how well their characteristics match. In this example we use the number of scenic cluster points within a 100 meter buffer of each computed route as the scenic measure under consideration. Table provides the descriptive statistics of scenic cluster points along individual routes in choice sets that were generated through the different algorithms described earlier. The bottom row shows these statistics also for chosen scenic routes. The results suggest that the maximum number of scenic points along a route (second column) as well as the mean number of scenic points along a route (fourth column) are similar between routes generated by the proposed branch-and-bound method and the observed scenic routes, which provides a better match than any of the other evaluated route generation methods. This shows that branch-and-bound algorithms can optimize route alternatives with regard to certain attributes. The large range of scenic points along generated routes (Max – Min) shows that this algorithm results also in a large variation between alternatives in a choice set along that attribute.

Another advantage of the proposed branch-and-bound algorithm is that it is not purely based on a cost function (between the origin and destination), but that it constrains the route to pass through a series of scenic points. Therefore, some of the generated alternatives based on this algorithm contain dead-end segments, realistic loops, and detours towards scenic spots, such as a viewpoint, which are also observed in VGI-based scenic routes. Such travel behavior cannot be replicated by labeling, link elimination or link penalty path generation methods. The generated route shown in Figure c provides an example of such a route with a realistic loop.

SUMMARY AND CONCLUSIONS
The computation of scenic routes requires knowledge about the attributes and features of the surrounding environment that make a route being perceived scenic. This has so far been only sparsely addressed in previous research. Route choice models provide an established framework to examine how travelers choose a scenic route and what attributes are significant in their perspective to choose one. This study proposed three algorithms (reduced link cost, reduced link cost with link elimination, branch-and-bound) to generate choice sets for a scenic route choice model. The characteristics of different algorithms in replicating chosen routes were examined
using the coverage measure and consistency index. The results show that the proposed branch-and-bound algorithm reproduces routes with better coverage rates as traditional deterministic approaches, whereas the proposed reduced link cost method as well as the reduced link cost with link elimination method did not show significant improvements. The branch-and-bound algorithm provides also a wider range in potential scenery related attributes (e.g. photo hot spots along a route) for route alternatives, which will improve model estimation performance. In addition, the branch-and-bound algorithm has the ability to generate routes that replicate the observed behavior of travelers in choosing routes with dead-ends. For future work we plan to further extend current research on scenic route choice modeling by including the presented expanded choice sets, which will allow estimating the importance of loops and dead-ends from a traveler’s perception of route scenery.

Acknowledgements
The authors are grateful to Google for financial support through a Google Research Award.

References


List of Figures
Figure 1 Computing a potential path area (PPA)
Figure 2 The branch-and-bound algorithm for the computation of scenic path alternatives
Figure 3 Computation of an alternative scenic route using the branch-and-bound algorithm
Figure 4 Analyzing detour and loops

List of Tables
Table 1 Coverage results and consistency index of applied algorithms at different thresholds
Table 2 Descriptive statistics for the number of scenic points along individual routes for different choice sets
a. Service areas for origin and destination for three travel-time combinations

Intersection

b. Intersection of two created service areas

Union

c. Final potential path area (PPA) for the specified origin and destination

Figure 1 Computing a potential path area (PPA)
Figure 2 The branch-and-bound algorithm for the computation of scenic path alternatives
Figure 3 Computation of an alternative scenic route using the branch-and-bound algorithm
a. Generated path with a long detour

Direct distance from start point to end point = 349 m
Direct distance from start point to scenic point = 95 m

SF = 158.25

b. An undesired loop caused in the path due to the selection of a scenic point

Direct distance from start point to end point = 0 m
Direct distance from start point to scenic point = 476 m

SF = 15.69  SF = 176.63  SF = 101.75

Loop 1  Loop 2

SF = 15.69  SF = 176.63

Loop 3

SF = 101.75

Loop 2

SF = 158.25

End point

Start point

Loop 3

c. Three nested loops that are created due to the selection of a scenic point

Figure 4 Analyzing detour and loops
Table 1 Coverage results and consistency index of applied algorithms at different thresholds

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Number of alternatives per O-D pair</th>
<th>Number of unique routes</th>
<th>Overlap threshold</th>
<th>Consistency index</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>90%</td>
<td>80%</td>
</tr>
<tr>
<td>Labeling approach (fastest route)</td>
<td>1</td>
<td>96</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Labeling approach (shortest route)</td>
<td>1</td>
<td>96</td>
<td>0</td>
<td>1.04</td>
</tr>
<tr>
<td>Labeling approach (minimum highway)</td>
<td>1</td>
<td>96</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Labeling approach (simplest route)</td>
<td>1</td>
<td>96</td>
<td>3.13</td>
<td>5.21</td>
</tr>
<tr>
<td>Labeling approach (minimum road change)</td>
<td>1</td>
<td>96</td>
<td>1.04</td>
<td>2.08</td>
</tr>
<tr>
<td>Link elimination approach</td>
<td>Max 10</td>
<td>940</td>
<td>5.21</td>
<td>10.42</td>
</tr>
<tr>
<td>Link penalty approach</td>
<td>15</td>
<td>1354</td>
<td>7.03</td>
<td>20.83</td>
</tr>
<tr>
<td>Labeling approach (reduced link cost)</td>
<td>1</td>
<td>96</td>
<td>2.08</td>
<td>3.13</td>
</tr>
<tr>
<td>Reduced cost-link elimination</td>
<td>Max 50</td>
<td>2282</td>
<td>11.7</td>
<td>20.21</td>
</tr>
<tr>
<td>Constrained enumeration (branch-and-bound)</td>
<td>Max 50</td>
<td>2241</td>
<td>15.54</td>
<td>24.88</td>
</tr>
<tr>
<td><strong>Merged choice set</strong></td>
<td>Max 131</td>
<td>7393</td>
<td>28.04</td>
<td>42.58</td>
</tr>
<tr>
<td>Algorithm</td>
<td>Min</td>
<td>Max</td>
<td>Median</td>
<td>Mean</td>
</tr>
<tr>
<td>-----------------------------------------------</td>
<td>-----</td>
<td>-----</td>
<td>--------</td>
<td>-------</td>
</tr>
<tr>
<td>Labeling approach (fastest route)</td>
<td>0</td>
<td>67</td>
<td>4.5</td>
<td>8.22</td>
</tr>
<tr>
<td>Labeling approach (shortest route)</td>
<td>0</td>
<td>81</td>
<td>4.5</td>
<td>9.07</td>
</tr>
<tr>
<td>Labeling approach (minimum highway)</td>
<td>0</td>
<td>67</td>
<td>4.5</td>
<td>8.17</td>
</tr>
<tr>
<td>Labeling approach (simplest route)</td>
<td>0</td>
<td>83</td>
<td>5</td>
<td>10.43</td>
</tr>
<tr>
<td>Labeling approach (minimum road change)</td>
<td>0</td>
<td>81</td>
<td>5</td>
<td>10.39</td>
</tr>
<tr>
<td>Link elimination approach</td>
<td>0</td>
<td>82</td>
<td>7</td>
<td>11.01</td>
</tr>
<tr>
<td>Link penalty approach</td>
<td>0</td>
<td>81</td>
<td>6</td>
<td>9.66</td>
</tr>
<tr>
<td>Labeling approach (reduced link cost)</td>
<td>0</td>
<td>66</td>
<td>5.5</td>
<td>10.62</td>
</tr>
<tr>
<td>Reduced cost-link elimination</td>
<td>0</td>
<td>101</td>
<td>10</td>
<td>14.31</td>
</tr>
<tr>
<td>Constrained enumeration (branch-and-bound)</td>
<td>1</td>
<td>104</td>
<td>15</td>
<td>19.72</td>
</tr>
<tr>
<td>Chosen routes</td>
<td>0</td>
<td>105</td>
<td>10</td>
<td>19.91</td>
</tr>
</tbody>
</table>