Archetypes of Alternative Routes in Buildings

Sebastian Feld, Martin Werner, Mirco Schönfeld, Stefanie Hasler
Mobile and Distributed Systems Group
Ludwig-Maximilians-Universität München
sebastian.feld@ifi.lmu.de, martin.werner@ifi.lmu.de,
mirco.schoenfeld@ifi.lmu.de, stefanie.hasler@gmail.com

***Abstract*—Alternative routes have found many applications in navigation scenarios. However, alternative routes have only been introduced recently for the indoor space due to the complexity of these environments. Furthermore, the number of alternative routes in buildings can be quite high. With this paper, we propose to organize sets of alternative routes by employing archetypal analysis on a feature space representation of routes and show results in which a set of hundreds of routes between the same start and end point has been compressed to only a few obviously different archetypal routes. Additionally, the framework allows for comparing routes with archetypes and with each other. This comparison does not reveal spatial similarity alone, but rather a measure of routes' similarity representing their inherent semantic character.**

I. Introduction

Navigation is surely one of the most frequently used applications with mobile devices. Even for indoor scenarios there is a constantly growing quantity of use cases. Think of construction workers that have to inspect several machines in complex industrial buildings, visitors of unknown premises like hospitals, museums or airports, and mobile robots in store houses collecting goods [25].

An extension to the classical wayfinding problem is the identification of alternative routes. This topic is handled well for outdoor scenarios like street networks, see for example the formidable survey of Bast et al. [3]. But there is also an increasing need for alternative routes in indoor scenarios. Think, for example, of firefighters needing an alternative to a given yet blocked route, or a navigation system at an airport proposing different routes in order to proactively prevent congestions or as a basis for multi-criteria optimization along different paths (e.g., types of shops).

Like just stated, there is much literature on the identification of alternative routes in street networks [1], [10], [21]. Basically, the algorithms focus on finding routes that differ on the highways mainly used. Unfortunately, the preconditions in indoor scenarios are quite different. The main limitation is the higher degree of freedom of movement as compared to street networks: a person can walk almost freely inside the corridors and halls resulting in possible turns not only at crossroads. The first definition of alternative routes in indoor scenarios has been given by Werner and Feld [28]. In summary, they define two routes having the same start and end point as proper alternatives if they traverse obstacles like walls or pillars on different sides. See Figure 1 as an example showing four alternative routes.

978-1-4673-8402-5/15/$31.00

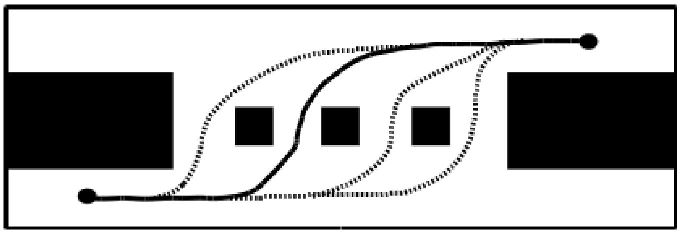


Fig. 1. An example showing four alternative routes. However, the dotted lines can be regarded as variations of the solid route.

The strength of this approach is the simple and clear definition together with the fact that this idea results in an equivalence relation. The main drawback is that the definitions given in [28] quickly lead to rather large sets of alternative routes in noisy floorplans, since small artefacts like furniture lead to the identification of alternative routes, even if the routes just have small variations.

With this paper, we concentrate on the question how to extract small sets of alternative routes with pairwise sensible dissimilarity from the set of all given alternative routes between two points. Note that the following analysis has been done with alternative routes in mind, but can be applied to any set of routes, for example when tracking multiple mobile devices in a building. Therefore, we propose to utilize archetypal analysis [8] – a statistical method for analyzing multivariate data sets – in the field of indoor navigation and in particular as a postprocessing step for the further understanding of given alternative routes. Given a floorplan and a (large) set of routes having the same start and end point, the objective is to find a small subset such that these routes are "pure types" (called archetypes), i.e. they represent ideal observations the other data points are combinations of. Now, the selection of routes will no longer focus only on the geometry or shape of the route (like the homotopy-based approach of [28]), but additionally on their particular nature and properties defined by the archetypes. In summary, we cluster a given set of routes based on their similarity to extreme examples called archetypes.

The main contributions of this paper are: (1) The definition of *abstract archetypal routes*, *realized archetypal routes*, and *faithful archetypal routes*. (2) The definition of a novel measure for route similarity, the *archetypal distance* between routes. (3) A framework to postprocess a given set of alternative routes in order to filter, analyze, and interpret them for a better understanding of the relation between routes and map.

The paper is structured as follows: Section II reviews related work in the fields of alternative routes, route similarity, and archetypal analysis. Section III introduces our concept of archetypal routes and archetypal distance followed by Section IV that describes our framework for calculating and analyzing archetypal routes. We provide a detailed evaluation of our concepts in Section V and conclude the paper in Section VI.

II. Related Work

This section discusses related work in the fields of alternative routes, route similarity, and archetypal analysis.

A. Alternative Routes

Basically, the problem of finding a shortest path between two points on a map can be modeled as a problem of wayfinding on a weighted graph. There are several approaches to solve this problem, with Dijkstra's algorithm [11] and A* [17] the most prominent ones.

The task of finding alternative routes in outdoor scenarios, i.e. on street networks, is discussed quite much in literature [3]. A notable example is the Penalty algorithm [21] that iterates between calculating a shortest path and increasing the corresponding edge weights such that eventually the shortest path might change. The Plateau approach [1] performs a forward and backward search in parallel creating candidates out of the preferably high overlappings of the searches. Finally, there are algorithms as described in [10] that try to create a set of multiple Pareto-maximal paths. Pareto-maximal paths with respect to a set of features are paths that are not dominated by others which means, that there are no paths that are at least equally good in all features and strictly better in at least one feature. Features of a trajectory in this context include length, time, costs, or number of turns.

Despite the fact that there is an increasing need for proper alternative routes in indoor scenarios, there is not much work in this field. Furthermore, the concepts and algorithms focussing on street networks cannot be mapped unmodified onto indoor scenarios. The main restraints are that usually there are no different types of ground floor like highways or small roads, and there is a much higher degree of freedom since a person can walk almost freely performing turns not only at crossroads.

The first definition of alternative routes in indoor navigation scenarios has been given by Werner and Feld [28]. They propose to use the topological concept of homotopy [4] in order to differentiate between equivalent and alternative routes. Basically, two routes having the same start and end point are regarded as proper alternatives if they traverse an obstacle on different sides. In a slightly more formal fashion: two routes p and q are alternative to each other if they are non-homotopic with respect to each other ($p \not\simeq q$). This relation can be approximated (under some simplifications ignoring winding numbers) by the question, whether the polygon spanned up contains at least one obstacle:

$$p \not\simeq q \Leftrightarrow \mathrm{polygon}(p * q^{-1}) \cap \mathrm{canvas} \neq \emptyset$$

with $*$ denoting the concatenation of compatible path segments and q^{-1} the inverse path of q. If, however, the routes p and q are homotopic to each other ($p \simeq q$), since they traverse obstacles in the same manner, the routes are regarded as equivalent. See [28] for more details on this definition.

B. Route Similarity

The calculation of alternative routes is closely related to the measurement of distances and similarity between routes: we would like to have preferably *different* routes. While it is very easy to assign a reasonable distance to points in geometry, it is a surprisingly complex topic when turning to trajectories. One reason for that is that the set of all routes (e.g., the set of all continuous maps from $[0, 1]$ into some metric space) is an infinite-dimensional space of functions leading to infinitely many reasonable ways of assigning a distance to two such objects. There is a plethora of algorithms that calculates route similarity with different strengths and weaknesses with respect to given requirements, but also in terms of complexity, runtime, and sensitivity to outlier and sampling rates [27, Chap. 6].

In general, most such distances of trajectories are expressed as algorithms incorporating the distance of points and the minimal distance between points and higher-dimensional objects such as line segments, trajectories or general point sets.

A family of very simple algorithms tries to represent the distance of trajectories as a summary over the distances of supporting points, especially for piecewise linear representations. These include the *Closest Pair Distance* [7] that calculates the distance between all points and returns the minimal distance, and the *Sum of Pair Distance* [2] that calculates the sum of the distances of all pairs of points. Both algorithms are prone to outliers, additionally the latter just works with trajectories having the same length. This restriction can, however, be overcome by inserting samples into both trajectories by interpolation until both trajectories contain the same number of points. As a consequence, this distance is not very efficient as the number of points that might be added is in the order of the length of both trajectories. In the worst case, for every point of the first trajectory, a new point in the second trajectory is introduced and vice versa. As the algorithms of this class are quadratic in the number of points, this is hardly acceptable.

Another family of algorithms tries to simplify the comparison of trajectories by adapting well-known distance approaches for strings: find common subtrajectories or edit one trajectory into the other. The method *Longest Common Subsequence* [26] originates from speech recognition and gives the length of the longest common subsequence. The ability to ignore very distant points makes this algorithm quite robust to noise and outliers. The *Edit Distance on Real Subsequences* [6], [22] counts the number of insert, delete, and replace operations of points that are needed to transform one trajectory into the other. *Dynamic Time Warping* (DTW) is a method that calculates the distance between two trajectories by minimizing the sum of the distances of pairs of points over a specific class of matchings of points. It is also known as the “frog distance” as it is the shortest distance a frog jumping forth and back between both sequences having only the choice to jump back to the next spot or the same spot on each trajectory has to take.

A third family of algorithms is purely motivated from geometry and reduces the complexity by effectively removing aspects of the trajectory's time domain. The *Hausdorff Distance* [19] is often used to calculate the similarity of point sets and it computes the maximum of the minimal distances between the two trajectories. As such, it only incorporates the spatial points and their distances and ignores the time dimension completely. Consequently, it can be very efficient and useful, but from time to time also misleading. A well-known refinement of the Hausdorff distance is given by the *Fréchet Distance* [15]. This distance is defined as an optimization over all possible monotonic reparametrizations (e.g., changing the flow of the time variable, but never the direction). It is best described as the minimal length of a leash connecting a dog and his owner while both are moving on two trajectories, but never backwards. It is noteworthy, that the optimization problem can be solved efficiently for piecewise linear trajectories (e.g., polylines) by iteratively solving a (finite) set of decision problems of whether a leash of a given length would suffice. While this approach is actually correct, it is also computationally expensive. Still, the discrete Fréchet distance, which limits the discussion only to the distances between points (e.g., find the shortest leash that could connect all pairs of points in a monotonous walk) has a known and quite small error bound. This enables the use of the Fréchet distance for pairwise comparison of trajectories at least for medium-sized collections of trajectories.

Even though the algorithms just mentioned are used in a wide variety of use cases, they are not appropriate for our scenario of alternative routes inside buildings, since they are – in general – unable to incorporate the underlying floorplan. However, there were proposed extensions that do involve the geometry of the map like the *Homotopic Fréchet Distance* [5], but those algorithms can not be applied towards analyzing large sets of routes (such as alternative routes) due to their immense running time.

C. Archetypal Analysis

Archetypal analysis [8] is a technique for statistical data analysis. It yields results that are comparable to those of clustering methods such as k-means clustering [18], for example. The goal of clustering is to separate data distributed over a feature space into useful partitions. There is a host of clustering algorithms that can achieve this. As an example, k-means clustering works by extracting medoids centered on data and aggregating the other data points to those so-called cluster centers.

Archetypal analysis also separates certain amounts of not necessarily equally spaced data. As opposed to traditional clustering methods, archetypal analysis looks for the points on the outer rim of the data space, approximating the convex hull of the data. In other words, it searches data points that are maximally distinct from each other.

The base algorithm as described in [8] and [24] is an iterative one which alternates between two steps. The goal is to find an approximation of the convex hull of the data space using comparatively few points. To that end, a linear optimization problem has to be solved.

Consider a data set with N observations (in our case alternative routes) and m attributes (e.g. length, number of turns, etc.) that is represented by an $N \times m$ matrix X. To extract a given number of k archetypes, the algorithm tries to find the $k \times m$-dimensional matrix Z by minimizing the residual sum of squares (RSS)

$$RSS = \|X - \alpha Z^T\|_2 \tag{1}$$

Thus, matrix X is compared to the product of the $N \times k$-dimensional coefficient α and the matrix of archetypes Z. $\|\cdot\|_2$ represents a fitting matrix norm, in this case the L_2-norm. With other words, α is the coefficient matrix needed to generate X from a given set of archetypes Z.

With the first iteration step the algorithm tries to minimize Equation (1) considering the constraints

$$\alpha_{ij} \geq 0 \text{ and } \sum_{j=1}^{k} \alpha_{ij} = 1$$

for $i = 1, ..., N$. The matrix of archetypes Z is a convex combination of real data points, that means it can be represented by

$$Z = X^T \beta \tag{2}$$

with β as an $N \times k$-dimensional matrix. With the second iteration step the algorithm tries to estimate Equation (2) considering the constraints

$$\beta_{ji} \geq 0 \text{ and } \sum_{i=1}^{N} \beta_{ji} = 1$$

for $j = 1, ..., k$. In a nutshell: The approach described in [8] is also known as alternating least square algorithm since it alternates between calculating the best coefficient α for given archetypes Z and calculating the best archetypes Z for given coefficient α.

Archetypal analysis iterates until it finds a minimum. It always terminates, but does not necessarily find the global minimum of the RSS (i.e., the best approximation of the convex hull of the data using k points), instead yielding a local minimum. Furthermore, there is no universal rule for the initial determination k, the number of archetypes. One commonly used approach is the "elbow" criterion, where a flattening of the screeplot of the RSS indicates a possibly good value of k. Further literature on details like numerical issues, stability, computational complexity, robustness, and concrete applications can be found in [8], [13], [14], [24].

III. Archetypal Routes and Distance

This section describes our concepts of archetypal routes as well as archetypal distance.

A. Archetypes of Alternative Routes

Archetypal analysis is about approximating the convex hull of observations in a multidimensional feature space with preferably few points. These points that define the approximated convex hull are called archetypes, i.e. they are not necessarily observed, extreme data points that describe the given data set well.

Like mentioned in Section II-C, the observations can be represented as convex combinations of the archetypes, and the archetypes can be constructed by convex combinations of the observations. This calculation takes place in the feature space, i.e. each observation and each archetype will be represented by a concrete configuration of the feature set.

One of this paper's main contributions is the definition of three different archetypal routes: the *abstract* archetypal route, the *realized* archetypal route, and the *faithful* archetypal route.

Abstract archetypal route: We call a concrete configuration of a feature set an abstract archetypal route. This point in the multidimensional feature space is not necessarily observed, actually it is very rare in our scenario of routes inside buildings. Note that abstract archetypal routes are equivalent to the original description of archetypes given in [8]. Figure 3, that will be explained in Section V-B, shows three abstract archetypal routes with their concrete feature configurations.

Realized archetypal route: Archetypes (our nomenclature: abstract archetypal routes) are points in a multidimensional feature space, that approximate the convex hull around the data points. The observations are points in this feature space as well. Thus, we can determine a concrete representative for each archetype using a certain algorithm, i.e. we "realize" the archetypes. The realized archetypal routes will then be concrete observations inside the instance space. With this paper we propose to use a simple "nearest neighbor" algorithm regarding the values of the coefficient matrix α. Like described in Section II-C, archetypal analysis describes each observation using coefficient matrix α. We now define the realized archetypal route to be the real observation, for which the coefficients of the representation as a sum of archetypes (as given by the matrix α) contains a maximum value with respect to the appropriate column. See Figure 4 for a representation of the observations inside the feature space. The point that is the nearest to an archetype will thus be chosen as the realized archetypal route.

Faithful archetypal route: We call concrete observations "faithful archetypal routes" if they have got exactly the same configuration of the feature set like a calculated abstract archetypal route. This can be realized archetypal routes with an appropriate α value of 1, but in the most cases this will certainly be synthesized data points. The question of how to create routes having predefined characteristics is very interesting and surely hard, and will be left open for future work. These class of faithful archetypal routes is especially interesting as the error introduced by realizing a given abstract archetype by a nearest neighbor as described in the previous section would be zero. With other words: if there exists a realized archetypal route (a concrete observation) that equals the abstract archetypal route, then it is also a faithful archetypal route. Otherwise, the data set does not contain a faithful archetypal route and it would have to be synthesized.

B. Archetypal Distance

Using archetypal analysis, one can summarize a dataset by a set of abstract archetypes which essentially are feature configurations of extreme cases. The feature space allows for direct calculation of similarities between observations and archetypes by means of the Euclidean norm of the difference of features.

Consequently, one can define the archetypal distance between two trajectories as the distance in feature space inbetween both trajectories. This also results in a distance between two archetypes and can be used to align embedded archetypes on various geometries. In order to visualize the results of an archetypal analysis, it is, for example, possible to form a distance matrix for the feature vectors of all archetypes and all real observations and to use multidimensional scaling [9] in order to embed these points in a way such that similarity and dissimilarity are preserved as good as possible.

In consequence, one can refine the archetypal distance to be a Euclidean distance of observations after embedding observations into a lower-dimensional Euclidean space using, for example, multi-dimensional scaling techniques. This is especially interesting, when novel observations, which were not part of the archetypal analysis, have to be introduced. These can easily be mapped into this low-dimensional embedding by means of lateration.

To make the definition of archetypal distance more handy: If one accepts the realization of archetypal routes to be a set of routes instead of a single route, one can also realize the routes using the top x observations or the observations having appropriate α values greater than a given threshold. This is exactly what archetypal distance means: The set of realized archetypal routes have got a small archetypal distance to each other.

IV. Framework for Calculating Archetypal Routes

This section describes our framework for calculating archetypal routes including the employed features that we use as a proof of concept for our definitions of archetypal routes and archetypal distance.

A. Concept and Implementation

We have implemented the framework completely in R [20], a software environment to perform and visualize statistical calculations.

Input of our algorithm is a floorplan of a building given as a common bitmap with white pixels representing walkable space and black pixels denoting obstacles like walls or furniture.

Further input is a "route store" consisting of tupels $< x, y, route\_id, class\_id >$ that where created using the indoor penalty algorithm proposed in [28]. x and y represent the coordinates of a point of a route. $route\_id$ assigns points to a route while $class\_id$ assigns a route to an equivalence class defined using the topological concept of homotopy (see Section II-A). Two routes having the same $class\_id$ are homotopic to each other and can be seen as equivalent, whereas two routes having different values of $class\_id$ are non-homotopic to each other and thus can be regarded as proper alternative routes.

The routes' features described in the following subsection are calculated using an own module called "feature extractor". This module is implemented solely using simple standard tools of R or, in the case of DTW, an existing library available via CRAN [16].

The archetypal analysis by itself is conducted using the R package "archetypes" [13], also available from CRAN, that was also used to create most of this paper's figures.

B. Selected Features

In order to apply archetypal analysis in the context of trajectory computing and alternative routes, we have to select a sensible set of features with which the routes can be represented as a numerical vector. By thorough experimentations, we found out that the following straigthforward numerical attributes of complete routes have a sufficient descriptive power. Still, the list is neither complete nor can be transferred to any domain or map without inspection. This, however, is a consequence of the impossibility of sensibly representing routes by small vectors of real numbers. Thus, the set of features used in this paper is just a proposal for a well-working feature set created by very simple means.

A concrete feature is described as a standardized floating point number representing an aspect of the route, e.g. its length. For each alternative route the same set of descriptive features is generated. These features are stored in a m-dimensional vector. These column vectors constitute the matrix X in the archetypal analysis (see Section II-C).

Basically, we have identified three groups of features: features regarding the geometric shape of a route, features regarding the relation to the shortest route, and features regarding the relation to the map.

1) Features regarding the shape:

chull_area: The area of the convex hull of all points of the route.

chull_size: The number of points that define the convex hull of all points of the route.

chull_centroid_x/y: The x and y coordinates of the centroid of the polygon generated by connecting all points of the route.

length: The absolute length of the route.

angularsum_cancelling/positive: The sum of the values of all turning angles. Since the routes are given by pixel-wise coordinates, the turning angles exist in 45 degree steps ranging from -180 to $+180$ (cancelling) or from 0 to 360 (positive).

2) Features regarding the shortest route:

relative_length: The length of the route divided by the length of the shortest route.

dtw: This feature calculates the distance of the given route to the shortest route using the dynamic time warping technique.

3) Features regarding the map:

average/min_heat: This feature uses a byproduct that is generated when calculating the set of alternative routes using the indoor penalty method described in [28]. The heat in a specific point on a map is defined as the number of routes of the penalty run that passed that spot. This heat can be normalized to a specific range or clipped to a certain interval. The average heat of a route is therefore the sum of the heat of each point on the route divided by the route's length.

V. Evaluation

This section contains elaborate experiments using different floorplans and it proceeds along the basic flow of archetypal analysis together with the discussion of particular phenomena.

A. Experiment Setup

We demonstrate and discuss our results using alternative routes calculated with four different floorplans. See Table I for a summary. Map "Office" is a very regular floorplan having rectangular rooms and corridors. Map "Spa" is a foorplan that is quite round and has a focus on its center. Map "Doom" is a simplified version of a map used in a first-person shooter game, having multiple irregular ways to follow. Finally, map "White House" is a simplified version of a historical floorplan of the White House.

TABLE I. Summary of the experiments

Map	*Dimension (Pixel)*	*Routes*	*Classes*
Office	1000 x 311	400	23
Spa	500 x 340	400	12
Doom	999 x 796	400	10
White House	529 x 361	400	37

Column *Routes* shows that we have calculated 400 routes using the framework proposed in [28] for every map. Like mentioned before, each route is assigned to an equivalence class defined by the homotopy relation. Thus, for example, the 400 routes of scenario "Office" are distributed over 23 homotopy classes (column *Classes*). Note that the number of routes and the number of homotopy classes depend on each other and that there potentially exist even more homotopy classes. More iterations of the penalty algorithm imply more routes imply potentially more homotopy classes.

For each map we have conducted multiple run of archetypal analysis with different values of k ranging from 1 to 10 in order to find and discuss different numbers of archetypes. We also repeated each experiment multiple times in order to prevent local minima. The archetypal analysis always converged and the results were reproducible.

B. Calculation of Abstract Archetypes

We would like to start the explanation of our framework's functioning as well as the evaluation of the results using map "Office". Archetypal analysis approximates the convex hull of the observations; thus, a good way to find the "correct" number of archetypes k is to inspect the residual sum of squares (RSS). A flattening of the curve indicates an appropriate value for k, since the additional archetype does not help in reducing the approximation error very much. This method is also called "elbow criterion", see [12], for example.

Figure 2 shows the RSS for different values of k. It is obvious that the RSS drops quite well from $k = 1$ to $k = 2$ and $k = 3$. But the difference from $k = 3$ to $k = 4$ is just marginal. Thus, we choose to fix $k = 3$ and further inspect the best model (since we have multiple iterations in order to avoid a local minima).

The best model for k archetypes can be well represented using barplots that show the feature configurations of the

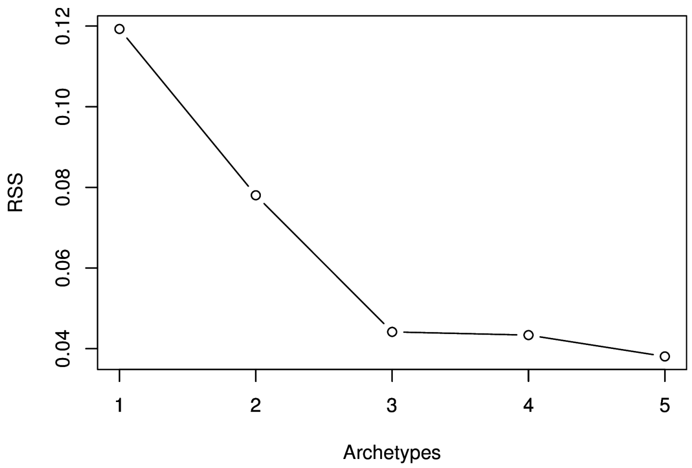


Fig. 2. Screeplot showing the resulting RSS for different values of k with map "Office". There is an "elbow" at $k = 3$.

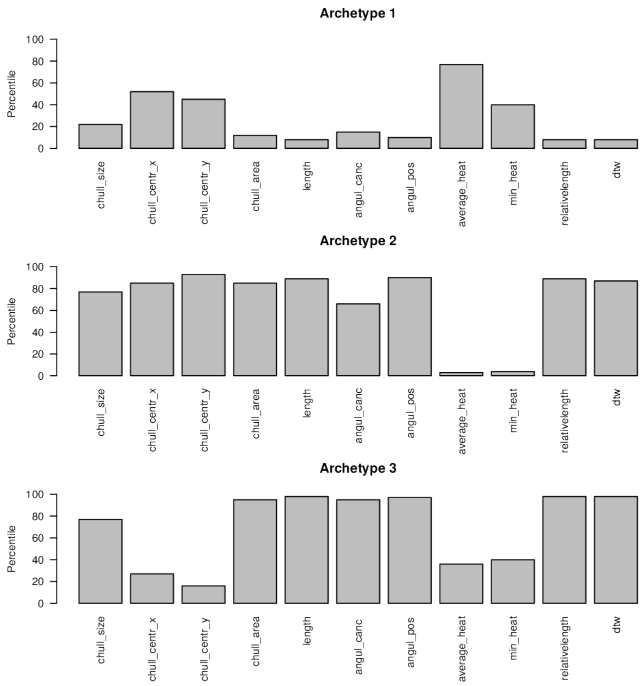


Fig. 3. Barplots for $k = 3$ representing the three abstract archetypal routes for map "Office".

archetypes (see Figure 3). These are what we consider the abstract archetypal routes.

After calculating a set of archetypes, there is the need of interpreting the results. Like stated before, map "Office" is a quite regular floorplan with rectangular rooms and corridors. Archetype $A1$, shown in the top row of Figure 3, has got low values for the convex hull's size (number of points) and area. That means that a route corresponding to archetype $A1$ is straight and strict, and goes more or less in a line-of-sight from start to goal. The moderate values for the x and y coordinates of the convex hull's centroid indicate that the main part of such a route traverses the map quite centrally. Low values for the absolute length, the length with respect to the shortest route,

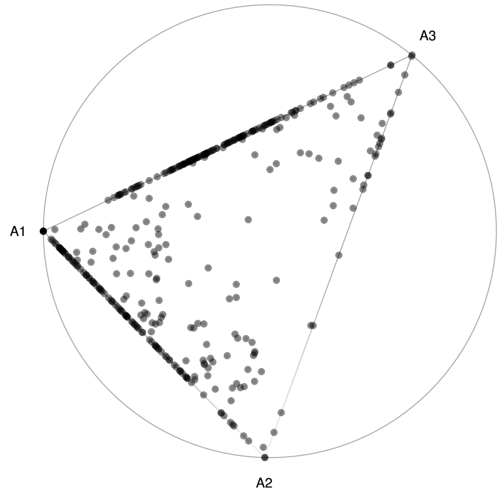


Fig. 4. Simplex plot showing the distances between the archetypes and the observations for $k = 3$ with map "Office".

the values for the angular sum, and the DTW distance to the shortest route indicate that a route corresponding to archetype $A1$ will be short by itself. Archetype $A2$ (middle row of Figure 3) has got high values for every feature except the heat. This means, that a corresponding route will be more like a detour; it is quite long, has got much turns and it traverses the map in the bottom part (point $(0, 0)$ is in the top-left corner of the floorplan). Archetype $A3$ has got high values for every feature, except quite low values for the convex hull's centroid and medium values for the heat. This means that a corresponding route traverses the map potentially in the upper-left part of the map and is quite long.

C. Realized Archetypes via Nearest Neighbor

A question that immediately follows is how the observations fit to the calculated archetypes. As explained in Section II-C, observations are convex compositions of the different archetypes (via coefficient matrix α).

Figure 4 shows a simplex plot with the archetypes at the corners of the triangle as well as the observations together with their pairwise distances. As one can see, there are observations that are very near to a concrete archetype. This brings us from abstract archetypal routes (points in the feature space) to the realized archetypal routes (concrete observations, thus: routes). As explained in Section III-A, we now identify the "nearest neighbor" of an abstract archetypal route with respect to the corresponding α value, thus we choose a concrete route to represent the realized archetypal route.

Now we can compare the resulting realized archetypal routes for $k = 3$ in Figure 5 with the barplot representation of the abstract archetypal routes in Figure 3. It is obvious that the interpretation and the chosen representatives for each archetype fit quite well. The strict and straight route in black corresponds to archetype $A1$, the long route at the bottom part of the map (red) represents archetype $A2$ and the green route $A3$. This filtering of three routes out of 400 is a first nice result, since the displayed routes can be regarded as pairwise different and surely appropriate for the use case of alternative routes in indoor navigation scenarios.

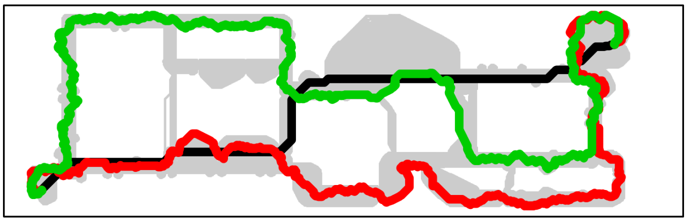


Fig. 5. Three realized archetypal routes for $k = 3$ in map "Office" (colored and bold). The gray routes in the background are the complete set of 400 routes given as input. Note that walls and obstacles are omitted in the figure for a better inspection.

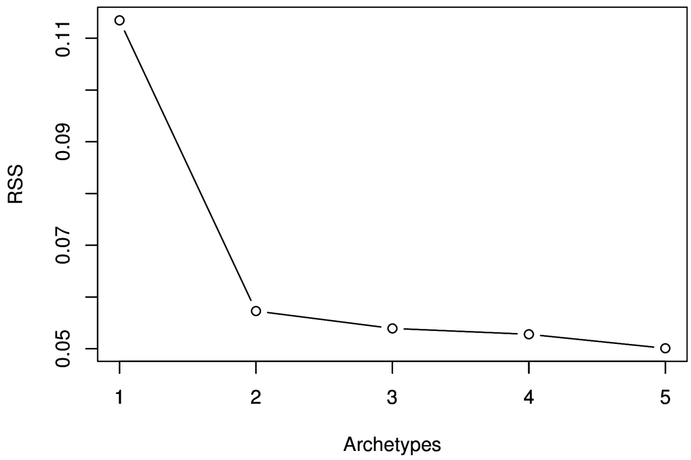


Fig. 6. Screeplot showing the resulting RSS for different values of k with map "Spa".

D. Relative Share vs. "Elbow Criterion"

Many papers in the domain of archetypal analysis propose to use the "elbow criterion" in combination with the screeplot of the RSS in order to fix the value for k, the number of archetypes. So did we with map "Office". In the course of our experiments it got evident that the additional focus on the relative distribution of the values of α may help finding an appropriate value for k. This section may demonstrate it with map "Spa".

Figure 6 shows the screeplot of the RSS for map "Spa". It is obvious that there is an "elbow" at $k = 2$, meaning that the reduction of the approximation's error is relatively small when adding another archetype. Figure 7 shows boxplots of the distribution of relative shares of the α values in order to get a second view. The diagram on the left-hand side of Figure 7 confirms that $k = 2$ was a good choice: the maximum values of both archetypes are extremely high (1 and 0.9996, respectively), and the upper quartile is at a value around 0.7. But in contrast to the screeplot, $k = 3$ is also a quite good choice when looking at the boxplot in the middle part of Figure 7. All three archetypes have got a very high maximum value (0.9993, 0.9991, and 0.9996, respectively), archetype $A1$ has got an upper quartile at 0.64, while the upper quartile for $A2$ and $A3$ is around 0.45. The boxplot for $k = 4$ is quite different. It is obvious, that the archetypes $A2$, $A3$, and $A4$ are defined or generated by outliers.

The next step would be to verify these thoughts by having a look at the realized archetypal routes, see Figure 8.

When describing the two realized archetypal routes for $k = 2$ (left-hand side of Figure 8) in a nutshell, one can see a very straight route traversing the lower part of the map and a quite winding route at the upper part. When choosing $k = 3$ (Figure 8, middle), there are two routes having the same properties but additionally a new one traversing the map's center. Please note that it is not necessarily the case that two archetypes stay more or less the same and a new one enters, since solutions are not nested with varying values of k [23]. The right-hand side of Figure 8 shows the realized archetypal routes for $k = 4$. This result is different than the other two, since we have three routes $A1$, $A2$, and $A4$ like with the setting $k = 3$, but additionally a new archetypal route $A3$ that is on the first glimpse quite identical to $A4$. In fact, the realized archetypal routes for $A3$ and $A4$ are homotopic to each other, i.e. they share the same homotopy class.

The results from above showed that we might get non-alternative, canonical routes for bigger values of k what led us to a further idea. Instead of limiting the number of k one can just react to the multiple appearances of homotopy classes. With other words: if there are routes having the same homotopy class but are assigned to different archetypes, choose the shortest of these routes and remove the other ones.

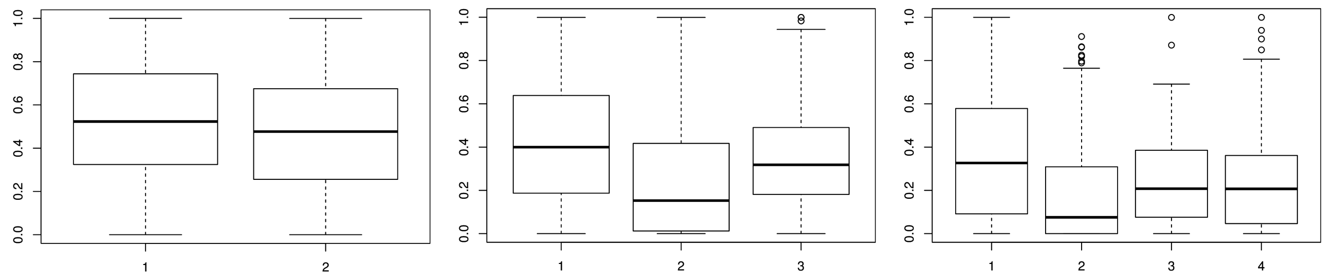
E. Archetypal Distance

The archetypal distance has been introduced in Section III-B. Now we would like to show the behavior of routes that are assigned to a concrete archetype to other routes. In particular: We expect that routes having a small archetypal distance to a concrete archetype (and thus, also to each other) show certain similarities. And at the same time we expect that routes having a large archetypal distance (they are assigned to different archetypes) show dissimilarities.

In order to examine this assumption we employed map "Doom" and created a set of "top routes" with respect to a certain archetype, i.e. we defined all observations having an appropriate α value greater than a fixed threshold to be part of this set. In this evaluation we set the threshold to be 0.8.

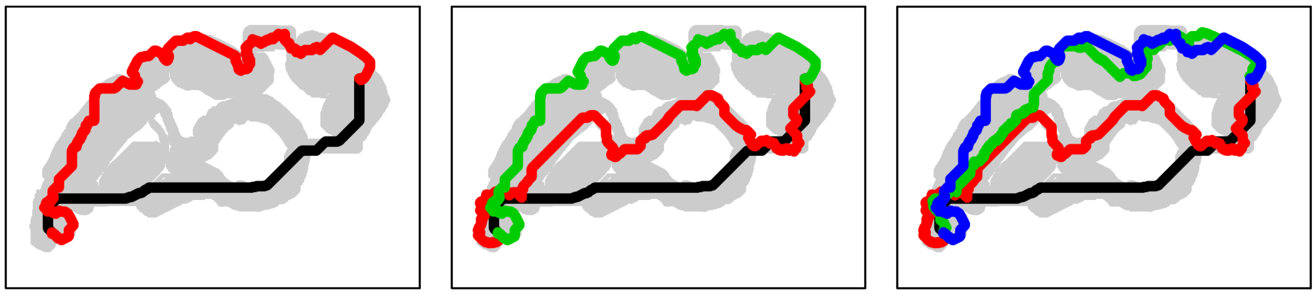
Figure 9 clearly shows that the set of "top routes" are similar to each other and show dissimilarities to the other archetypes, not only by means of "heading left, middle, right". The routes of archetype $A1$ are very straight and fast routes that traverse the map on the left-hand side. However, the routes assigned to archetype $A2$ are at the left-hand side as well, but are more winding than the former ones and they traverse the pillar at the top of the map just at the left-hand side. The routes of archetype $A3$ are quite straight and traverse the map at the right-hand side, while the routes assigned to archetype $A4$ head through the center of the map. Note that the set of "top routes" have different sizes. This shows that different archetypes may represent the given set of observations different well.

Figure 10 depicts the "top routes" of map "Office" that was investigated in a previous section. The figure shows clearly that the set of "top routes" are different to the realized archetypal routes (see Figure 5), but at the same time quite similar. The realized archetypal route for archetype $A1$ (black route in Figure 5) was described as "the strict one" going virtually line-of-sight from start to goal. Now, the set of "top routes" differs from that, since they basically cover the whole map.



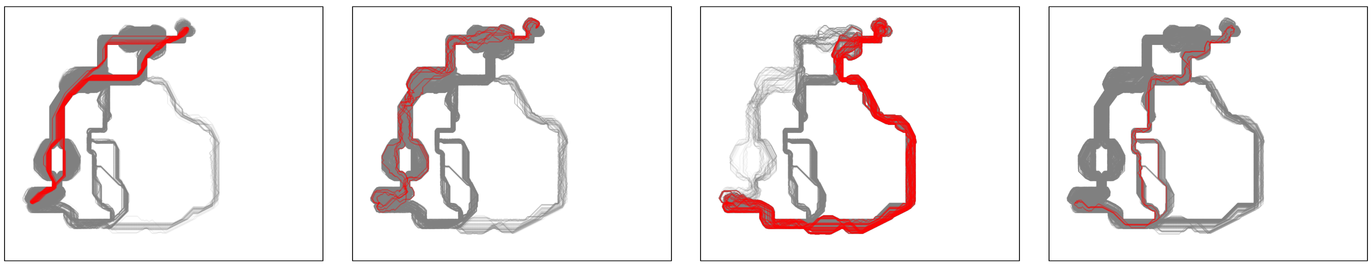
(a) $k = 2$ (b) $k = 3$ (c) $k = 4$

Fig. 7. Boxplot showing the distribution of relative shares of the α values for map “Spa”.



(a) $k = 2$ (b) $k = 3$ (c) $k = 4$

Fig. 8. Realized archetypal routes for different values of k in map “Spa” (colored and bold). Like in Figure 5 we plotted the complete set of routes in gray and omitted walls and obstacles for a better inspection.



(a) Archetype 1 of 4 (b) Archetype 2 of 4 (c) Archetype 3 of 4 (d) Archetype 4 of 4

Fig. 9. “Top routes” for each of the four archetypes in map “Doom”, i.e. the set of routes having an appropriate α value greater than a threshold of 0.8.

But the main characteristic, the strict and straight form, is still visible. The routes show just few turns. The realized archetypal route for $A2$ (red route in Figure 5) was described as winding and curly, but always following the lower part of the map. This is basically the same for the set of the “top routes”, just some small variations were added. The realized route for $A3$ (green route in Figure 5) was very long and can be described as a complete detour traversing the top and the bottom part of the map. Thus, it is a preferably long and winding route. This is again visible in the set of “top routes”, they show a wide variety of routes that are preferably long.

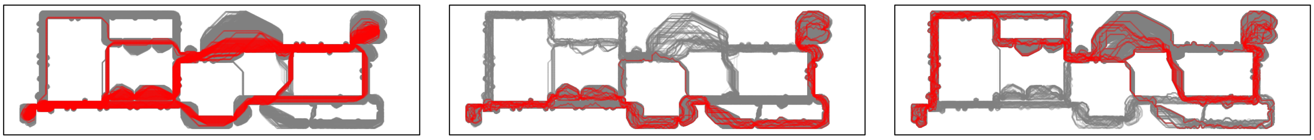
F. Homotopy Classes

Finally, we would like to discuss the aggregation power of the “top routes” regarding the homotopy classes of the given alternative routes. See Figure 11 showing the realized archetypal routes for different values of k in the scenario “White House”.

As with the experiments before, we calculated the set of “top routes” using a threshold of 0.8 for the α values. When choosing $k = 2$, the set of 400 routes gets reduced to 139 while assigning 87 routes out of 10 homotopy classes to archetype $A1$ and 52 routes out of 20 homotopy classes to archetype $A2$. When choosing $k = 3$, we assigned 42 routes out of 8 homotopy classes to archetype $A1$, 24 routes out of 7 homotopy classes to archetype $A2$, and 7 routes out of 4 homotopy classes to archetype $A3$. Finally, when choosing $k = 4$, we combined 27 routes out of 7 classes ($A1$), 4 routes out of 4 classes ($A2$), 5 routes out of 4 classes ($A3$), and 29 routes out of 9 classes ($A4$). Summarized, the classification considered many routes from many different homotopy classes.

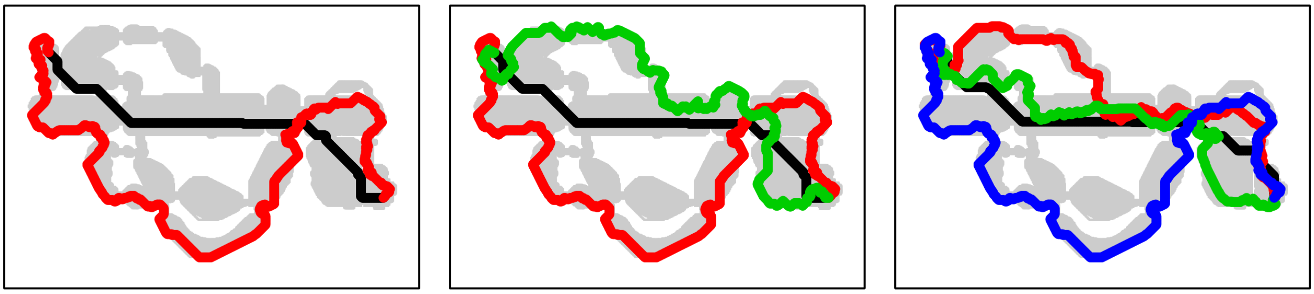
VI. Conclusion

With this paper we proposed to use archetypal analysis – a statistical method for analyzing multivariate data sets – in the field of indoor navigation and in particular as a postprocessing step for the further understanding of given alternative routes. We defined abstract, realized, and faithful archetypal routes and also a novel measure for route similarity, namely the



(a) Archetype 1 of 3 (b) Archetype 2 of 3 (c) Archetype 3 of 3

Fig. 10. "Top routes" for each of the three archetypes in map "Office", i.e. the set of routes having an appropriate α value greater than a threshold of 0.8.



(a) $k = 2$ (b) $k = 3$ (c) $k = 4$

Fig. 11. Realized archetypal routes for different values of k in map "White House" (colored and bold).

archetypal distance between routes. The archetypal distance helps to investigate given routes not only on the geometry or the shape, but additionally on the particular nature and properties defined by the archetypes.

With our implemented framework we have also shown that a rather simple set of features describing spatial trajectories is sufficient to postprocess a given set of routes in order to filter, analyze, and interpret them for a better understanding. Organizing these sets further allows for implementing complex multi-criterial optimization applications. We also performed thorough evaluations using different floorplans and scenarios, and human inspection on a vast amount of results.

For future work we envision to evaluate more and different features in this framework including their derivation, product and the like in order to add or remove certain features. Furthermore we would like to focus on the question of how to synthesize faithful archetypal routes. Finally a more detailed analysis of the archetypal distance is eligible regarding the behavior of concrete features from different abstract archetypes.

REFERENCES

[1] Cambridge Vehicle Information Technology Ltd. - Choice Routing. Online, 2013. http://www.camvit.com/camvit-technical-english/Camvit-Choice-Routing-Explanation-english.pdf.

[2] Rakesh Agrawal, Christos Faloutsos, and Arun Swami. Efficient similarity search in sequence databases. In David B. Lomet, editor, *Foundations of Data Organization and Algorithms*, volume 730 of *Lecture Notes in Computer Science*, pages 69–84. Springer Berlin Heidelberg, 1993.

[3] Hannah Bast, Daniel Delling, Andrew Goldberg, Matthias Müller-Hannemann, Thomas Pajor, Peter Sanders, Dorothea Wagner, and Renato Werneck. Route planning in transportation networks. In *Technical Report MSR-TR-2014-4*. Microsoft Research, Microsoft Corporation, 2014.

[4] Glen E Bredon. *Topology and geometry*, volume 139. Springer, 1993.

[5] Erin Wolf Chambers, Eric Colin De Verdiere, Jeff Erickson, Sylvain Lazard, Francis Lazarus, and Shripad Thite. Homotopic fréchet distance between curves or, walking your dog in the woods in polynomial time. *Computational Geometry*, 43(3):295–311, 2010.

[6] Lei Chen, M. Tamer Özsu, and Vincent Oria. Robust and fast similarity search for moving object trajectories. In *Proceedings of the 2005 ACM SIGMOD International Conference on Management of Data*, SIGMOD '05, pages 459–502, New York, NY, USA, 2005. ACM.

[7] Antonio Corral, Yannis Manolopoulos, Yannis Theodoridis, and Michael Vassilakopoulos. Closest pair queries in spatial databases. *SIGMOD Rec.*, 29(2), Mai 2000.

[8] Adele Cutler and Leo Breiman. Archetypal analysis. *Technometrics*, 36(4):338–347, 1994.

[9] Jan De Leeuw and Patrick Mair. Multidimensional scaling using majorization: Smacof in r. *Department of Statistics, UCLA*, 2011.

[10] Daniel Delling and Dorothea Wagner. Pareto paths with sharc. In *Experimental Algorithms*, pages 125–136. Springer, 2009.

[11] E. W. Dijkstra. A note on two problems in connexion with graphs. *Numerische Mathematik*, 1(1):269–271, 1959.

[12] Manuel J. A. Eugster. Archetypal athletes, 2011.

[13] Manuel J. A. Eugster and Friedrich Leisch. From spider-man to hero – archetypal analysis in r. *Journal of Statistical Software*, 30(8):1–23, 2009.

[14] Manuel J. A. Eugster and Friedrich Leisch. Weighted and robust archetypal analysis. *Computational Statistics & Data Analysis*, 55(3):1215–1225, 2011.

[15] Maurice Fréchet. Sur quelques points du calcul fonctionnel. *Rendiconti del Circolo Matematico di Palermo*, 22(1):1–72, 1906.

[16] Toni Giorgino. Computing and visualizing dynamic time warping alignments in r: the dtw package. *Journal of statistical Software*, 31(7):1–24, 2009.

[17] Peter E Hart, Nils J Nilsson, and Bertram Raphael. A formal basis for the heuristic determination of minimum cost paths. *Systems Science and Cybernetics, IEEE Transactions on*, 4(2):100–107, 1968.

[18] John A. Hartigan. *Clustering Algorithms*. John Wiley & Sons, Inc., 1975.

[19] Felix Hausdorff. *Grundzüge der Mengenlehre*. Verlag von Veit und Comp., 1914.

[20] Ross Ihaka and Robert Gentleman. R: a language for data analysis and graphics. *Journal of computational and graphical statistics*, 5(3):299–314, 1996.

[21] Moritz Kobitzsch, Marcel Radermacher, and Dennis Schieferdecker. Evolution and evaluation of the penalty method for alternative graphs. In *Proceedings of the 13th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS'13)*, pages 94–107, 2013.

[22] Vladimir Iossifowitsch Levenshtein. Binary codes capable of correcting deletions, insertions and reversals. In *Soviet physics doklady*, volume 10, page 707, 1966.

[23] Giovanni C Porzio, Giancarlo Ragozini, and Domenico Vistocco. On the use of archetypes as benchmarks. *Applied Stochastic Models in Business and Industry*, 24(5):419–437, 2008.

[24] Christian Seiler and Klaus Wohlrabe. Archetypal scientists. CESifo Working Paper Series 3990, CESifo Group Munich, 2012.

[25] Sebastian Thrun. Learning metric-topological maps for indoor mobile robot navigation. *Artificial Intelligence*, 99(1):21–71, 1998.

[26] Michaeil Vlachos, Dimitrios Gunopoulos, and George Kollios. Discovering similar multidimensional trajectories. In *Proceedings of the 18th International Conference on Data Engineering*, ICDE '02, pages 673–684, Washington, DC, USA, 2002. IEEE Computer Society.

[27] Martin Werner. *Indoor Location-Based Services*. Springer, 2014.

[28] Martin Werner and Sebastian Feld. Homotopy and alternative routes in indoor navigation scenarios. In *Proceedings of the 5th International Conference on Indoor Positioning and Indoor Navigation (IPIN'14)*, 2014.