Road lighting efficiency improvement through data processing and automated computation

J. Starczewski
jak.starczewski@gmail.com
Instituto Superior Técnico, Universidade de Lisboa, Portugal
October 2020

***Abstract*—Road lighting design requires extraction of crucial parameters i.e. lighting situations from the data (both road and lighting infrastructure) regarding the area intended for the retrofit. As the commercially available software is not capable of batch processing geospatial data, this process is currently done by hand, which inevitably leads to many simplifications and omissions, especially in large-scale projects. For that reason, an attempt to automatically derive the lighting situations from the official data concerning the area of Washington, D.C has been performed. The process assumed determination of spatial relationships between three sets of data (street lights, street segments, and roads) in order to obtain the luminaires and their corresponding road segments. The procedures, by means of which the best ascription candidate was chosen, were formed by initially reviewing the problematic cases manually and extending the algorithms accordingly, until the satisfactory outcomes were achieved. Therefore, a set of criteria has been determined to minimize the ambiguity of the assignments. As a result, 99% of street lights have been automatically assigned to the road segments. The results suggest that the lighting design performed with the help of the developed solution yields significantly more accurate outcomes in comparison with the other considered approaches. As the tool is versatile, it also allows to automatically estimate the possible power savings for any other area (assuming similar data) without the need for human assistance. This is especially helpful during the strategic planning phase of the lighting retrofit projects.**

***Index Terms*—Street lighting, lighting design, energy efficiency, data processing, automated computation.**

I. Theoretical background

Investments in efficient outdoor lighting systems will result in immense energy savings, but the decision to upgrade the lighting infrastructure has to be carefully thought through and be made upon a thorough inference process. Fortunately, the investment potential can be found by collecting and analyzing the following sets of the data:

- Infrastructure and road configuration data,
- Data necessary for lighting class selection process [1]

GIS data considering lighting infrastructure and road layout can be collected in different forms including CAD files/drawings, Excel sheets, JSON and XML files or various GIS formats (ESRI Shapefiles, GeoJSON, GeoPackages, etc.). Similarly, data collected from either traditional or mobile sensing devices, complemented by freely available data from world maps like OSM, can support road network analysis and as a result – the selection of lighting classes for the examined segments.

The aforementioned sets of data serve as a basis for the most important part of every street lighting modernization project – the design phase itself. There are several objectives that the street light system designer has to accomplish. First and foremost, outdoor lighting design has to ensure safety and comfort of moving in public spaces, by providing adequate illumination during night-time i.e. by selecting the appropriate fixture models and their parameters. The selection is being made by means of photometric computations. The draft has to comply with current standards while maintaining cost-effectiveness, which can be represented by any quantitative criterion like the reduction of energy consumption or investment expenditure. Finally, the aspect of aesthetics also has to be taken into account [2]. Three feasible retrofit scopes based on lighting design paradigm are being distinguished:

- luminaire replacement - which is the most obvious modernization model. Using LED light sources instead of high-pressure sodium or metal halide luminaires, results in 40% energy usage reduction on average,
- lighting design optimization - which assumes reducing the power consumption to the lowest possible level, while at the same time conforming to the requirements of the current lighting standards,
- implementation of street lighting control - which is intended to select the necessary amount of light to be provided for each assessed road segment, proportionally to fluctuations in vehicle flow intensity, level of ambient light or weather conditions.

Their applicability is dependent on numerous factors like modernization type or financial restrictions etc. [1].

Nevertheless, the collected GIS data which is intended for the design procedure requires preprocessing, which involves the extraction of applicable calculation fields in accordance with CEN 13201-3 standard [3]. This is possible only after the identification of relevant areas together with their types (roads, sidewalks etc.) and luminaires associated with them. This, however, is an extremely demanding task for a computing system, specifically for projects covering whole cities [4].

Typically, each street is assessed by lighting engineers with regard to all parameters which affect lighting, such as pole distribution, pole spacing, road width, pole setbacks, etc. Nevertheless, each lighting situation needs to be configured separately by hand during this process. As tools such as DIALux do not analyse the input setting in any way, they expect the lighting situation to be already determined and described by an engineer. Programs dealing with photometric calculations do not recognise the geometric object along which the street lights are arranged, what they see, however, is a lighting situation with the poles configured in a specific distance between them. Therefore, as the manual examination of all lamp settings is not feasible in a timely manner, engineers tend to reduce the number of modelled situations, thus reducing precision and risking over and under-lighting.

On the other hand, there are also GIS-like systems, which are being used for continuous operational lamp management (e.g. Streetlight Inventory [5]). In these systems, lamps actually appear alongside the roads; however, only the layer of street lights is modelled. It is possible to view and check the parameters of particular lamps in these systems, however, as there is no representation of the roads as such, associating the points with the areas they illuminate is usually impossible. Street lights may even have a street name assigned to them (to ease the management); however, most of the time there is either no reference to the street, or the reference is too general, and thus, it cannot be concluded that the lamps are actually positioned alongside a geometric object. As the map visualizations underneath are based on one of the web mapping services available, the system is not aware of the streets which are there, and therefore it is inaccurate to claim that they light up a certain area.

These systems are excellent as GIS management tools, however, the problem is that they only model the lamp data. While they do it with exceptional precision, they do not understand the situations that they model, i.e. they have neither the possibility of manually modelling the lighting situations as there is no representation of illuminated area nor do they have any possibility of automating the process.

For the above reasons, the processing of the road segments is currently made by a human. The manual approach is understandable in case of deficient and rough data; however, considering the ubiquitous availability of structured and semantically described data, the attempt to automate the process is definitely worth pursuing, especially as the designer's time is generally much more valuable than that of the CPU. This is particularly helpful during the strategic planning phase, when the decision-makers need an accurate estimation of the possible power savings and investment costs, e.g. to prioritise the modernisation in different parts of the city.

II. Problem statement and proposed solution

The primary goal of the thesis is to investigate the possibility of utilizing existing GIS datasets to develop automatic algorithms to handle the tedious task of data integration and data analysis in road and street lighting modernization projects. This will be done by developing a tool to process geospatial data in order to automatically determine spatial relationships between the datasets.

The solution will be based on a set of procedures build upon the examination of the analyzed collections of data, starting from a manual review of questionable results and ending with fully automatic processing. The tool will encompass the following functionality targets:

- Cleaning, conversion, and unification of the data,
- Automatic assignment of segment linestrings to the lit areas,
- Automatic assignment of lamps to the lit areas,
- Evaluation of the results of the automatic process and road segments quality analysis,
- Preparation of the data for photometric optimization.

III. Road lighting efficiency improvement through data processing and automated computation

A. Selection of tools and libraries

During the planning phase of the thesis, a decision related to the environment of the experimental procedure development had to be made. The selection process was based on ease of use, efficiency and, most importantly, on the availability of the pre-existing packages within a certain programming language with the following characteristics:

- ability to read GIS vector data from different formats such as ESRI Shapefile, GeoPackage or GeoJSON,
- ability to deal with different projections of the data (transform the data from one projection to another),
- ability to clean and manipulate the data,
- ability to perform spatial operations on the geo-referenced data.

Due to the compliance with the mentioned requirements, the Python programming language has been chosen as the primary technology to be used in the approach. As there is a variety of open-source libraries designed to perform geospatial operations, Python emerges as a front-runner in GIS-related applications.

Three main Python packages will be utilized during the development of the proposed solution:

- pandas, which will be used to cleanse and manipulate the data,
- Shapely, which allows to perform numerous geospatial operations on a variety of geometries such as points, linestrings, and polygons,
- GeoPandas, which is built on top of the two preceding packages (extends their functionalities). GeoPandas supports many different input data formats and is equipped with a number of features for spatial analysis and processing, which significantly eases the work with geospatial data. GeoPandas is priceless as it comes to performing bulk operations on geometric attributes or during spatial joins.

B. Cleaning, conversion and unification of the data

The data is provided as datasets and may come in different forms, including Excel sheets, JSON and XML files, AutoCAD drawings, and various GIS formats (ESRI Shapefiles, GeoJSON, GeoPackages, etc.), and therefore, in case of incompatible data collections, the standardization process will be applied to ensure that data is consistent. Datasets will be read by the GeoPandas module and both their spatial data types (Points, LineStrings, Polygons) and coordinate reference systems will be analyzed. Imported data will have a certain spatial reference already assigned, however it is usually much more convenient to convert latitude and longitude values into projected coordinates, which enables efficient calculations, including distance. Precise selection of the projected coordinate system together with its unification within each dataset will be performed to ensure the accuracy of the results, as the inconsistencies in this aspect might become a reason for algorithm misbehavior. Then, after the identification of attributes and their semantics, unnecessary columns will be dropped from the datasets. Moreover, the data collections will be filtered out by the values in specific columns, which will not be useful in further calculations. Finally, a number of cleaning operations will be performed on all datasets, such as:

- Unifying remaining attributes' names and their values (e.g. transforming all string values to lowercase),
- Converting strings to numeric values for future calculations (e.g. light pole height),
- Converting units from imperial/US units to the metric system,
- Replacing various string representations of N/A values with consistent numpy equivalent (e.g. np.nan),
- Saving clean datasets in one common format

C. Automatic assignment of segment linestrings to the lit area shapes

Once the data is consistent and in a calculation-ready form, the process will be initialized by performing a spatial join operation on road polygons and segment linestrings. Spatial join combines the data in a similar fashion as the standard join, nonetheless, the spatial affinity of two features is the relation basis [6]. This relationship between geometric objects is defined by means of Shapely's binary predicates, which return a boolean value, as they are implemented as methods [7]. The following predicates can be used to determine how the spatial join operation will be utilized, by specifying the `op` parameter [8]:

- `within` – this predicate returns `True` if the first object (considering both its interior and boundary) intersects only with the interior of the other object,
- `contains` – the inverse of within predicate,
- `intersects` – the default spatial join setting, which returns `True` if the objects have common points [7]

The last predicate will be used in the spatial join, as all road polygons which intersect with a specific road segment linestring need to be acquired. Therefore, the attributes of the latter will be joined into the former, and as a result, each road polygon would get the attributes of a linestring that intersect with the polygon. This operation would enable the future calculations – the assignment of lamps to the lit areas – to be performed not only basing on the physical distance between the examined layers but also on other significant segment attributes like the street name or its type. Then, a DataFrame that consist of road segments and a set of intersecting road polygons will be created. The information about the segment's geometry will be merged into the established object and a GeoDataFrame with an appropriate coordinate reference system will be eventually constructed. At this point in the process, a buffer will be performed on the segment linestring, which is one of the most commonly used operations in spatial analytics. By applying a buffer to any of the geometries, a circumferential polygon is created around the original object within a certain distance [6].

Then, in order to derive road segments polygons, road geometries grouped by the linestring they intersect with, will undergo a union operation, which combines multiple shapes into a single object. It has to be noted, however, that from the acquired union polygon, only the parts which intersect with a certain segment's buffer will be taken into consideration for further examination (Fig. 1).

Fig. 1. The union of polygons (violet) intersecting with a specific segment's linestring buffer (blue)

D. Automatic assignment of lamps to the lit areas

Again, the process will be initialized by performing a spatial join operation, however, this time the computation will consider previously created road segment polygons and street light points. Before the operation, points that refer to the position of light poles will be buffered by a specified distance. By doing so, it will be possible to obtain several candidate roads assigned to each of the lighting points (2).

The spatially-joined table will be then grouped by the identification number of the light pole and, for each road segment in a group, a number of factors will be calculated, including the physical distance from the pole to road segment or Levenshtein distance between street names of both datasets.[1] Based on these values, a normalized score within the range of (0;1) will be obtained for each candidate road segment. The coefficients obtained from the following methods will determine how accurate each assignment is:

[1]Levenshtein distance is defined as a minimum number of character operations (e.g substitute, delete or insert) to transform an expression into another [9].

- distance score – which calculates a partial score based on physical distance from street light to the road segment.
- name score – which compares Levenshtein distances between street name attributes. Essentially, it splits the street names ascribed to both datasets into a list of strings and calculates the minimum distance in terms of similarity between any of the examined elements in the collection.
- type score – which evaluates the type of road the light is being ascribed to. It is based on the assumption that the lamps will most likely illuminate the streets of dominant categories (primary, residential etc.), while the others will receive a slightly lower rating (e.g. service road).
- number of lamps in a segment – which applies the partial score of 1 if the number of luminaires assigned to a given road segment is more than one, otherwise, the partial score of 0 is applied.

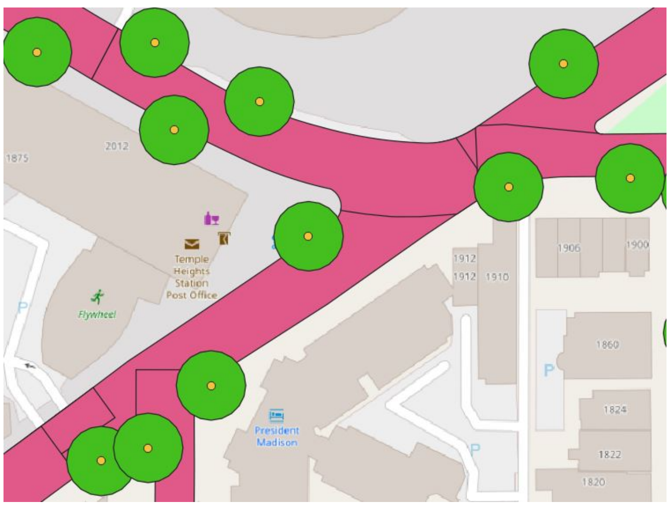

Fig. 2. Buffered street light points (green) together with the candidating road segments (violet)

Weights for each of the partial scores will be determined empirically, the overall score will be computed and the road segment with the highest overall score will be ascribed to the street light. The outcome of the automatic process will be evaluated by manually reviewing the cases. The examination of the results will then allow to introduce additional heuristics or adjust a number of parameters to improve the accuracy of the algorithm, for example:

- The distance parameter of the buffer operation – if the distance is too low, the buffer of the street light might not intersect with any of the road segments,
- The lowest acceptable score based on which the assignment is made,
- Weights of specific scores.

E. Road segment quality analysis

After the adjustment of the results, road segments will be analyzed by calculating and examining several road properties, which will serve as quality indicators of the modeled lighting situation. First of all, the width of the road from the perspective of each street light assigned to the segment will be acquired. For that purpose, a straight line will be constructed through the point which represents the position of the street light and the nearest point between the mentioned point and the road segment. The latter geometry will be obtained with the help of Shapely's `nearest_points` function, which calculates the closest points of two geometries. Again with the help of Shapely, the linestring will be extrapolated in both directions by means of a scale function, which is available from the affinity module. The intersection length between the linestring and road segment polygon will therefore characterize the width property.

Then, information related to the street segment linestring geometry will be merged to the main GeoDataFrame, which consist of street lights and associated road segment polygons, and basing on the spatial relationship between them, the spacing between the lamps in each road segment will be calculated. This indicator will be obtained in the following steps:

- First, points representing the street lights will be projected on the linestring which depicts the segment using Shapely's `project` method. As a result of this operation, the distance between the lamp and the beginning of the linestring will be obtained [7].
- Then, each group of lamps will be sorted by the acquired distance and the position of the lamp in a sequence will be defined.
- Finally, the spacing between each neighboring lamp in a sequence will be calculated. This operation will be performed with pandas' `diff` function, which computes the difference between two elements of the DataFrame [10] (in this particular case the difference between the value and the next element in the row)

Both indicators will undergo a statistical analysis, and after reviewing questionable outcomes, the results will be tuned.

F. Preparation of the data for photometric optimisation

The final results will be converted into the input format of PhoCa software, which will be used for the optimization of examined lighting installation. The platform requires each row of the dataset to be a distinct street light, with the following attributes: unique identifier of fixture-segment pair, street light identifier, street name, road segment identifier, number of fixtures, road width, number of lanes*, lighting class*, arrangement of the street lights (one-sided or double-sided)*, spacing of street lights in a segment, distance from the street light to the nearest edge of road segment, height of the light pole, arm length of the light pole, fixture type (road, park or decorative)*, fixture wattage, latitude, longitude. Attributes marked with an asterisk refer to the data, which is not available from within the available datasets, thus it will be derived analytically.

IV. Case Study

A. Dataset description

The study will cover the District of Columbia, with the data publicly available from Open Data DC[2] and GEOFABRIK[3] websites, with the datasets from the latter being based on the OpenStreetMap data. Both of them provide numerous spatial datasets for the analyzed area, available for download in several formats, including ESRI Shapefiles. Calculations will be performed on the following datasets:

- Street Lights (ESRI Shapefile) – containing street lights together with their locations and attributes (source: Open Data DC),
- Street Segments (ESRI Shapefile) – illustrated as linestrings representing the axis of particular roads in the District (source: GEOFABRIK),
- Roads (ESRI Shapefile) – which consists of precise roadway shapes in the form of polygons (source: Open Data DC)

1) Street lights: The streetlights dataset (70,956 records) is the fundamental dataset used for the analysis of the current installation. It models each lighting pole as a point with features, which characterize its geometry and additional parameters. The following data is available within the attributes of the dataset:

- Attributes correlated with the position of the light pole like the street name, ward, or the city quadrant, within which the records are located,
- Data concerning the US equivalent of European road lighting classes,
- A set of information regarding lighting poles like the pole height or its type,
- Data describing the style, length, and number of arms of a specific light pole,
- Fixture data, which consists of its type, style, number of lights together with their wattage,
- The most important attribute, which characterizes each record in the dataset geographically.

Figure 3 is a classification of poles concerning the number of fixtures installed (lighting poles with 1 fixture were ignored in the following figure, as they make up the vast majority of the dataset i.e. 67,705 records). As the figure 3 indicates, there are 3,251 lighting poles with more than 1 fixture. The dataset, however, does not provide the information about areas they illuminate, therefore is it important to determine a number of carriageways associated with each pole. Considering the multitude of street light types, it is important to categorize them by a number of areas they illuminate. This can be done by applying heuristics based on other features such as the arm lengths, fixture style etc. Therefore, to ensure the accuracy of lighting design, the following analysis has been performed:

- Street lights with two fixtures and number of arms equal 2 have been acquired (735 records).
- The observation of the records with double arms suggests that the shorter arm lights up pavement, and therefore lights with asymmetrical arm lengths has been filtered out (661 records remaining).
- Lights has been also analyzed in terms of `arm_style` attribute. Fixtures with the arm style of either ”Decorated Straight” or ”WP Decorated Scroll” have been identified to be positioned in parallel with the road and has been dropped consequently (507 records remaining).
- The aforementioned conclusion was also applicable for lights with `asset_type` of ”Bridge Light”, thus these 98 records have also been rejected.

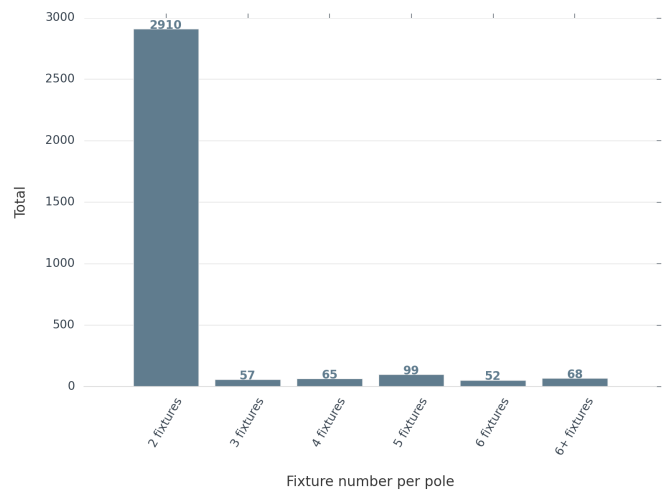

Fig. 3. Number of fixtures installed per pole

Eventually, 409 street lights, which illuminate 2 areas, have been determined. All lights that have been dropped during analysis, will be considered as a single lights in further calculations.

2) Street segments: Street segments (32,841 records) represents all streets as linestrings. The dataset contains, except the ever-present identification number and the geometry coordinates of the shape, several valuable attributes including the type of the road segment with the corresponding code, street name, highway location reference, maximum allowed speed and the street layer (useful for identification of e.g. underground roads).

3) Roads: The roads dataset (52,339 records) contains polygons, which represent the roads in the District of Columbia. The precise shapes have be obtained by combining LiDaR data with orthophoto imagery and other map sources. The records contain an information about their area type e.g. road, paved drive, intersection, median island, parking lot, etc. Other attributes in the dataset include identification numbers, the capture date, as well as the geographical features of the shape such as the area or length.

B. Cleaning, conversion and unification of the data

After reading the data with the GeoPandas module, the datasets have undergone an analysis of the coordinate refer-

[2]https://opendata.dc.gov
[3]https://geofabrik.de

ence system assigned to the geometry attribute. The spatial data has been transformed to a projected coordinate system (NAD83). Then, a number of cleaning procedures described in IV-B has been performed. Finally, filtering operations, described below, have been carried out individually on each dataset, and afterward, attributes which have not been perceived as particularly useful have been dropped.

- Street Lights (filtered using the `asset_type` attribute) – As the analysis scope will exclusively cover street lights along the main roads, only records with asset types of either 'Street Light' or 'Bridge Light' will be taken into account in further calculations.
- Street Segments (filtered using the `fclass`, `tunnel` and `layer` attributes) – Driven by the same reasoning, segments with 'fclass' values referring to one of the main carriageway types in Washington, D.C were preserved. Moreover, records that suggest underpass in the 'tunnel' attribute and values below 0 in the 'layer' feature have been omitted, as underground lighting will not participate in the computations.
- Roads (filtered using the `description` attribute) – Similarly to the aforementioned datasets, only records characterized by 'Road', 'Intersection' or 'Hidden Road' values in the description column have been preserved, with the latter retained to keep the consistency of several carriageways.

Cleaned datasets have been saved in GeoPackage format, which proved to be more flexible than Shapefiles e.g. for the visualization purposes.

C. Automatic assignment of segment linestrings to the lit area shapes

A spatial join performed in the first iteration of the process resulted in a GeoDataFrame consisting of 50,621 records. 5,679 segments and 140 roads have not intersected at all, which corresponds to 23% and almost 1% of the initial size of the datasets respectively. Then, by grouping the GeoDataFrame object by the `id` of a road segment and aggregating all identifiers of intersecting road polygons, the DataFrame with the row length of 18,983, has been created. After merging the segment's geometry information to the above table, a buffer of 10 meters has been applied to the object's spatial attribute. A distance of 10m has been selected to avoid shape protrusion which may affect the assignment process. Polygons in each group underwent the union operation and the shapes have been trimmed by previously retrieved linestring buffer. The distribution of the number of polygons intersecting with each street segment has been presented in Figure 4. The procedure was successful for all road segments, and as a consequence, a total of 18,998 road segments have been created. Most of the constructed objects have yielded a satisfactory result (Fig. 1), however, after a scrupulous review, a major issue has been identified. During the process, a significant number of multipolygons (2,930 out of 18,998) has been formed, which is undesirable for further calculations. Problematic geometries have been identified as originating from the following reasons:

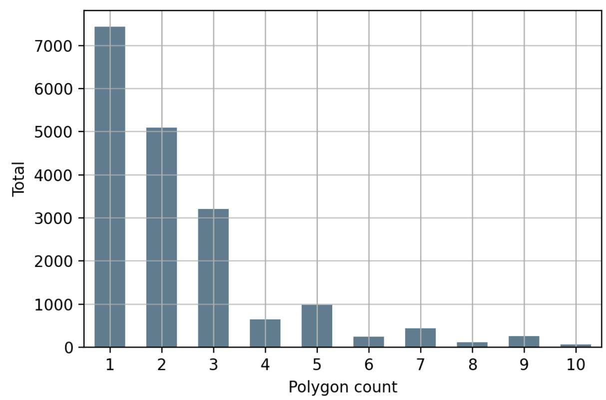

Fig. 4. The distribution of the number of polygons intersecting with each street segment (1-10 polygon groups only, as the larger groups constitute of only 2.5% of cases)

- segment intersects with roads only on its ends,
- segment crosses the road centrally through the previously filtered median island,
- occurrence of barely noticeable holes between road polygons,
- inaccurate road shapes.

In the second approach to the process of assigning segment linestrings to the lit areas, solutions to the aforementioned issues have been applied, which involved e.g. calculation of the total intersection percentage between linestring and road segment polygon. The issue with the creation of multipolygons has been resolved and the acquired results were satisfactory.

D. Automatic assignment of lamps to the lit areas

Similarly to IV-C, the procedure was started by executing a spatial join (again based on the intersection relationship) between previously created road segment polygons and buffered points, which refer to street lights (a buffer of 5m has been initially applied to the dataset containing street light data). The operation resulted in a GeoDataFrame consisting of 73,184 records with 648 non-intersecting lamps. By reviewing records that were not included in the geopandas data structure, it was found that the applied distance was to low, as several lamps, which in fact illuminate a certain area, has been disregarded. Therefore, spatial join has been conducted again with a higher distance of 7 meters, producing a new GeoDataFrame (79,743 records) which will be the basis for the further calculations. The assignment will be performed in accordance with the distinction made in IV-A1 (number of areas illuminated by street lights).

1) Assignment of lamps, which illuminate two areas: As described in IV-A1 a total of 409 lights with this characteristic have been identified recognized, which translates into 1,015 records in the GeoDataFrame resulting from the spatial join. 8 lamps, which do not intersect with any nearby area has been identified. Moreover, further 30 records, which intersect with one area only have been found. After manual examination of the aforementioned street lights, it was concluded that both records which were not included in the spatial join and records which did not intersect will the minimum of 2 areas, will be considered as double lights which light up a single segment. The following assignment accuracy measures will be applied for the remaining 371 fixtures and the candidate road segments:

- distance score – The setback of 0 to 3 meters is practically an equally good assignment and it is a matter of adjusting the mounting angle of a fixture by several degrees. At some point, as the distance is significantly greater, this possibility vanishes. Due to this fact, the normalization formula will be an exponential function, instead of standard linear dependency and has been described in (1).

$$distance\_score = 1 - (1/1 + e^{5-distance}) \qquad (1)$$

  The exponent in 1 has been also empirically tuned in accordance with the maximum distance between the lighting pole and road segment, which was formerly defined as 7 meters. Therefore, candidates with the setback of 0-3 meters will receive scores between 1 and 0.88 respectively, a setback of 5m will correspond to the score of 0.5, and road segments with a distance of around 7 meters will be assigned a score around 0.
- type score – the lamps, which illuminate the streets of dominant categories will receive a score of 1, otherwise a score of 0.8 will be applied (service road, unclassified road),
- intersection percentage – as the road segment polygons with low intersection percentage has not been omitted (to prove the accuracy of the algorithm and not to accidentally skip correct candidates), this measure will be additionally used to determine to most suitable assignment candidates.

Weights for each score have been set as equally important and the overall score has been computed. Then, the assignment process has been performed, taking into consideration only the candidate roads, with the overall score higher than 0.5. The results of the assignment have been satisfactory, and therefore the weights of particular scores have not been adjusted. However, several incorrect ascriptions have been spotted, thus the minimum score needed to be increased. In order to acquire an accurate minimum score, all assignments within the score of 0.5–0.7 have been examined. As a result, the minimum assignment score of 0.6 has been obtained. Again, street light with an insufficient score (3 records), will be considered as a single-area double lights in the further calculations.

2) Assignment of lamps which illuminate a single area: Initially, the already assigned street lights have been dropped from the GeoDataFrame (78,734 records remaining). 509 records have not been ascribed, however, after manual examination it was decided not to increase the maximum distance, as it would not increase the accuracy of the algorithm. The matter is much more related to the absence of data in certain areas (either road polygons or segment linestrings) or to the incorrect asset type in the initial dataset (e.g. alley lights marked as street lights). The same assignment scores will be applied for the candidate road segments in this case, together with a pair of additional measures, namely, ”number of lamps in a segment” and the ”name score”. Comparison result between street name attributes of the latter score will also be normalized exponentially, however, the score will be much more strict. Candidates with the Levenshtein distance of 0 and 1 will receive scores of 1 and 0.92 respectively, a distance of 2 will correspond to the score of 0.5, and road segments with the higher distance will be assigned a score around 0. The formula has been described in (2).

$$name\_score = 1 - (1/1 + e^{5-(2.5*min\_lv\_distance})} \qquad (2)$$

Weights for each score have been also initially set as equally important and the overall score has been computed. The assignment process has been performed, taking into consideration only the candidate roads, with the overall score higher than 0.5. The results of the assignment have been satisfactory in general; however, several ascriptions have not been made, which was found to be related to wrong weight distribution. Therefore, the weights of the both the ”number of lamps in a segment” and ”intersection percentage” scores have been halved. After the weight adjustment, the process has been repeated. While the importance tuning handled the issue of wrong assignments, the minimum overall score needed to be increased as some undesired ascriptions have been made. Therefore, all assignments within the score of 0.5-0.6 have been examined and as a result, the minimum assignment score of 0.547 has been obtained. Consequently, 26 records below this score have not been assigned (otherwise, they would be assigned to the wrong area). On the other hand, the remaining 54,117 has been assigned the most suitable candidate road segment. Furthermore, street lights, which light up two areas, have been duplicated and transformed into single-area lamps (each record has been split into 2 separate rows, with a unique id and corresponding attribute values such as arm length or fixture wattage) and have been merged to the data structure (a total of 54,853 records). Based on the road segment identification numbers assigned to street lights, a dataset of all ascribed road segments (8,259 records) has been also acquired for further calculations.

E. Road segments quality analysis

1) Width of the road segment: The width of the road from the perspective of each street light assigned to the segment has been calculated in accordance with the procedure described in IV-E. However, after examining the results, several cases with width over 20 meters have been identified, which is especially odd, as the linestring buffer (IV-C) had been previously specified to be 10 meters (theoretically disqualifying width values above 20m). Therefore, records with problematic width values have been visualized and the following reasons of incorrect measurement have been determined:

- Street light positioned diagonally or laterally in relation to the road segment,
- Inaccurate attribute values,
- Inaccurate differentiation of street lights, which illuminate two road segments.

Precise width values are essential, especially as the road width median of each segment is necessary during the preparation of the data for photometric optimization. Therefore, in order to acquire representative width values, the following procedure has been implemented:

- If there is only one street light in a segment and has a width value of over 20 meters assigned, the dataset width median (10.3 m) will be applied to the value instead,
- If there are two street lights in a segment and one of them has a width value of over 20 meters assigned, the value will be leveled with the width of the second street light,
- If there are more than two street light in a segment and more (or equal) than half of the street lights ascribed have a width value of over 20 meters assigned, the dataset width median will be applied to all values instead.

The final distribution of road width has been presented in 5

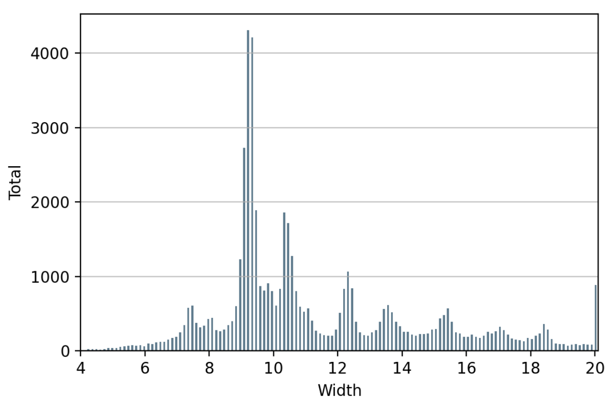

Fig. 5. Road width distribution

2) Spacing of lamps within a road segment: First of all, the spacing of street lights within each segment has been calculated in accordance with the procedure described in IV-E. Then, a median and a variance of spacing have been computed for each road segment, with the latter being used as a segment quality indicator. Generally - the lower the variance, the better. This is due to the fact that the median of spacing in a particular road segment is the calculation basis during the photometric optimization process, rather than its discrete spacing values and thus the outcomes are much more accurate. Spacing variance results have been analyzed and the following value groups have been distinguished:

- N/A value – 31.7% – Value associated with the pole count of less than 3 within a road segment. Segments with single street light do not have the spacing value at all and segments with two lights have only one, thus the variance is not computed in both cases. The lack of the indicator makes quality evaluation of these road segments impossible.
- Value lower than 100 – 48.3% – The analysis demonstrates, that low spacing discrepancy has been identified among the majority of examined road segments, which indicates their excellent quality.
- Value between 100 and 200 – 10.5% – Acceptable road segments with slight differences in spacing values.
- Value over 200 – 9.5% – Almost one-tenth of segments obtained a value of over 200, which suggests spacing issues within a segment. Therefore, records with inappropriate spacing variance have been visualized and the following reasons of excessive values have been determined: absence of a street light in a sequence, intersections disrupting spacing uniformity, bilateral arrangement. While the first two cases are self-explanatory, this particular situation requires clarification. The distribution of spacing within segments with the perfectly opposite arrangement is divided into two groups with regular values and figures around 0, which may result in ridiculously low spacing median values (Fig. 6).

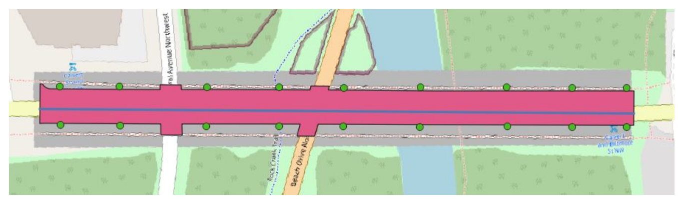

Fig. 6. Example of bilateral arrangement

As the automatic solutions to the first two problematic segment groups will be a part of the future work, these road segments, will feature in further calculations in their intact form. However, an arrangement of street lights in a segment will be calculated (unilateral/bilateral) to resolve the last issue. By providing the information about the bilateral arrangement to the software, it will automatically split the segment into two independent lighting situations, thus improving the accuracy of the optimization. The arrangement and adjusted spacing median will be obtained by implementing the following procedure to each road segment:

- Calculate the average segment spacing and split the set of the spacing values into two subsets (over and below average),
- Empirically acquire the maximum value of spacing, under which the segment might be classified as bilateral (3 meters),
- Calculate the ratio of spacing values lower than 3 meters in the collection that holds values below average,
- If the ratio is higher than 0.5, the road segment will be considered as bilateral. The ratio level of over 0.5 has been chosen as several cases have been observed, where the arrangement of the road segment could be considered bilateral, despite locating an occasional value higher than 3 meters in the collection,

- The spacing median of all records with detected bilaterality will be adjusted to the sum of the medians of values in groups over and below average (e.g. 20.5m + 0.5m, instead of 0.5m).

By applying the aforementioned methodology, 21% of road segments with problematic spacing variance have been corrected. Finally, the median results have also shown that 1,417 records (one lamp in a segment) have received an "N/A" value, thus the dataset spacing median (29.45 meters) has been assigned to these records instead.

F. Preparation of the data for photometric optimisation

Having calculated the arrangement of street lights within a segment in the previous subsection, the remaining required data, which is still not available is the number of lanes, road classification, and the type of fixture. The first mentioned information will be acquired by dividing the road median of each segment by 3.5, which is a common industry practice. As it comes to the road classification, it is one of the attributes of the initial street lights dataset, however, the volume of the pedestrian traffic(high, medium, low marked as "H", "M" and "L" respectively) is not defined, which is required in the U.S. standard. Therefore, as other sources are not accessible, the design will be performed basing on the classification obtained from custom conversion rules presented in I. The logic behind it is straightforward, i.e. the rules assume that the higher the theoretical pedestrian traffic volume of a particular road, the stricter the lighting requirements (e.g. "Local" road should have higher pedestrian traffic volume than the "Collector" road, therefore the former has been assigned a value of "H", while the latter a value of "M"). It has to be noted that "Arterial" roads have been classified as "MAJOR_A" and "Other Freeway and Expressway" value has been converted to "FREEWAY_A" as there was no possibility to differentiate between the types through other attributes.

TABLE I
CONVERSION OF ROAD CLASSIFICATION AVAILABLE IN THE DATASET TO THE OFFICIAL CLASSIFICATION VALUES ACCEPTED BY PHOCA SOFTWARE

Dataset Classification	Converted Classification
Local	LOCAL_H
Collector	COLLECTOR_M
Minor Arterial	MAJOR_L
Principal Arterial	MAJOR_L
Interstate	FREEWAY_A
Other Freeway and Expressway	FREEWAY_A

Last, but not least, the fixture type had to be decided between road, park, or decorative. It was assumed (and confirmed by visualization), that all records with "Posttop", "Mushroom" and "Chinatown Luminaire" values in the fixture style attribute will be assigned a fixture type of "park" (and "road" for the remaining street lights). The final results have been converted into the input format of PhoCa software and contained information about all assigned street lights (54,853 records).

V. RESULTS

The quality of the data preparation process in lighting retrofit projects can be evaluated with the number of extracted lighting situations and the extend to which they reflect the reality. Generally, the more configurations can be generated by a certain approach, the more precise the design will be and its representation of reality will be more accurate. Considering the available data, the following approaches to the retrofit of lighting installation in the District of Columbia are viable:

- luminaire replacement based on the conversion chart,
- luminaire replacement and photometric optimization based only on the initial data,
- luminaire replacement and photometric optimization based on the processed data, which underwent the procedures developed in the thesis.

The input format of the most photometric calculation software assumes that each row represents a fixture together with the assigned road segment parameters. However, as it would be inefficient to consider each luminaire separately, they group these situations by the parameters significant for the optimization process i.e. road width, spacing, pole height, arm length etc. Therefore, the aforementioned approaches will be assessed based on the number of groups created from the available parameters.

The first approach is burdened with immense error, as there is no guarantee regarding the quality of the existing installation - it may display severe under- or over-lighting. The number of different configurations is equal to the count of unique values representing energy expenditure in the `wattage` attribute. Based on these values, a LED fixture with an appropriately lower wattage is installed.

The second approach assumes utilization of photometric calculations rather than using a conversion chart. It is based on a notably greater number of parameters, as the initial dataset (besides the wattage data) contains the information about the height of the pole, arm length and fixture type. However, there is no data related to the distance from the pole to the edge of the road, road width or pole spacing. These values have to be estimated using assumptions based on the type of the road i.e. higher pole spacing on the freeways than on the residential roads, etc.

Finally, the third approach, based on all the aforementioned parameters together with the ones acquired via the developed tool. Unlike the previous procedure, it assumes either minimum or no estimations at all (considering width and spacing). All approaches have been grouped by the available parameters and the results have been presented in Fig. 7.

Despite the fact that the number of distinguished groups is incomparably greater in my approach, it is also based on values which are closer to the real situations. Even though the procedure developed in the thesis assumes some estimates, they are much more accurate in comparison to the other approaches. The proposed procedure allows to calculate lighting situations more precisely, and more importantly, without the involvement of human factor. The tool is not capable

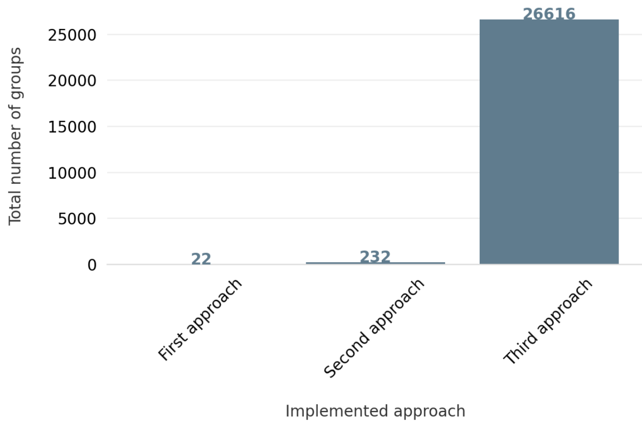

Fig. 7. A total number of configurations depending on the implemented approach

of performing a detailed design (at this point), however, it allows to precisely estimate the power required for lighting installation in a fully automatic manner. Not only is it reusable for other areas (assuming analogous sets of data), it is also much more accurate than the other approaches.

VI. Conclusions

Road lighting design accuracy depends on correct extraction and interpretation of lighting situations (parameters which take part in the photometric optimisation process) from the available data. Due to the enormous computation scale and the inability to batch process in most of industry-standard applications, the operation is currently done manually (which leads to many simplifications). Therefore, an attempt to automatically derive the lighting situations from the data concerning the District of Columbia has been performed. The presented methodology was based on the determination of spatial relationships between three sets of data, namely: street lights (which models each lighting pole as a point with additional features), street segments (which represent streets as linestrings), and roads (which refer to the shapes of the carriageways in a form of polygons). Initially, the precise road shapes have been associated with intersecting street segments and have been split in accordance with the outer points of the linestrings. The process was followed by the assignment of light points to the previously created road segments, thus developing a set of lighting situations. The calculations were carried out iteratively and the procedures were extended until satisfactory outcomes were achieved (99% of street lights assigned to the road segments). While the accuracy of the assignment cannot be evaluated precisely (as it would require manual review of more than 50,000 records), it has to be emphasized, that lighting design would be impossible to perform in a timely manner by means of the standard procedures for a project of this scale. Therefore, a thoughtful selection of ascription criteria together with careful adjustment of their corresponding weights, made it possible to obtain the most optimal results possible. These methodology assumptions minimized the ambiguity of the assignments, which was one of the successfully overcome challenges, The remaining problems, i.e. the identification of relevant data within the analyzed dataset and the differentiation of lamps which light up multiple areas have been also solved. With regard to the former, the relevant computation data (main carriageways in Washington, D.C.) have been obtained through analysis of attributes and visualisation of records. The latter, however, was managed by applying heuristics based on numerous attributes of street lights such as arm lengths and fixture style. Consequently, 409 records have been recognized to illuminate two distinct road segments.

Finally, as the results described in section V suggest, the lighting design performed with the help of the developed tool yields superior outcomes in comparison with the other considered approaches. Not only are they significantly more precise, they also reflect the reality in a much more accurate way. The number of possible configurations created by means of the proposed solution was respectively 1,210 and 114 times greater than by utilization of other procedures described in V. The tool allows to automatically (no human interference) estimate the possible power savings in a certain area (assuming similar data), which is especially desirable during the planning phase of the lighting retrofit projects.

References

1 Kotulski, L., Landtsheer, J. D., Penninck, S., Sedziwy, A., and Wojnicki, I., “Supporting energy efficiency optimization in lighting design process,” p. 12, 2013.

2 Sedziwy, A. and Kozień-Woźniak, M., “Computational Support for Optimizing Street Lighting Design,” in *Complex Systems and Dependability*, Zamojski, W., Mazurkiewicz, J., Sugier, J., Walkowiak, T., and Kacprzyk, J., Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, vol. 170, pp. 241–255, reporter: Complex Systems and Dependability Series Title: Advances in Intelligent and Soft Computing. [Online]. Available: http://link.springer.com/10.1007/978-3-642-30662-4_16

3 *EN 13201-3:2015 Road lighting - Part 3: Calculation of performance*. Warszawa: Polski Komitet Normalizacyjny, 2016, oCLC: 1150419442.

4 Sedziwy, A. and Kotulski, L., “Multi-agent System Supporting Automated GIS-based Photometric Computations,” *Procedia Computer Science*, vol. 80, pp. 824–833, 2016, reporter: Procedia Computer Science. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S187705091630847X

5 “Streetlight inventory.” [Online]. Available: https://solutions.arcgis.com/local-government/help/streetlight-inventory/

6 Lawhead, J., *Learning geospatial analysis with Python: understand GIS fundamentals and perform remote sensing data analysis using Python 3.7*, 2019, oCLC: 1154476837.

7 “Shapely.” [Online]. Available: https://shapely.readthedocs.io/en/latest/manual.html

8 “GeoPandas.” [Online]. Available: https://geopandas.org/index.html

9 “Levenshtein distance.” [Online]. Available: https://en.wikipedia.org/wiki/Levenshtein_distance

10 “pandas.” [Online]. Available: https://pandas.pydata.org/