479

Deep Learning Video Analytics to Assess VGA Measures and in Public Spaces

The case of a pedestrian public place: Piazza Duomo, Milan, Italy

FEDERICO MESSA, GIULIA CECCARELLI, ANDREA GORRINI, DANTE PRESICCE & RAWAD CHOUBASSI,

Fondazione Transform Transport ETS

ABSTRACT

Since the introduction of the Social Logic of Space and the further developments of Space Syntax theories during the following decades, the proposed methodologies have been proven effective to analyse the space in his physical configuration. In this regard, the study aims to contribute to this research topic, proposing the application of Visibility Graph Analysis (VGA) measures to Piazza Duomo, one of the busiest public squares in Milan, Italy. These metrics were tested in relation to the pedestrian movements measured through remote sensing technologies. In particular, deep-learning video-analytics methods were used to detect the pedestrian occupancy in five different moments of the day. Detections were then reprojected on the plane to obtain the spatial utilization, discretized on the same grid used to calculate the VGA metrics and a series of correlation analyses tested the grid-based analyses against the cumulated footfall and a movement data metric. The results show how at different radii, considered through the variation of the restricted visibility parameter, the correlation between the VGA measures and the pedestrian movements varies greatly, ranging from negative weak correlations to positive moderate correlations, in the case of cumulative density, and from positive moderate correlations to weak correlations in the case of the turnover metric. The results, in line with other recent research works, shows how pure VGA measures can only partially describe an open public space, posing an interesting step in the direction of future works, aimed at the development of a more complex model based on the Space Syntax theories.

KEYWORDS

Visibility Graph Analysis, Grid-based Analysis Video Analytics, Public Space, Pedestrian Counting

1 INTRODUCTION

Since the introduction of the Social Logic of Space (Hillier and Hanson 1984) and the further development of Space Syntax theories (Hillier et al. 1996, Hillier 2007), the proposed methodologies have been thoroughly studied in an effort to build a theoretical and practical framework to investigate and quantify the characteristics of the space. The geometrical and topological structure of the environment (Hillier 1999) and the way the users explore the space (Benedikt 1979) are the foundation of the Visibility Graph Analysis (VGA), which is the main focus of this research. The Grid-based analysis was firstly introduced as an undirected graph built on a set of regularly-spaced points placed in the "voids" of the built environment (Turner et al. 2001, Turner 2001) and was further developed (Turner 2003) to include many of the syntactic measures already tested on the Line-based methodologies (i.e., *axial maps*).

This research's scope is to use a set of the available VGA metrics to describe a pedestrian public space, specifically Piazza Duomo in Milan, Italy, and to confront the outcomes with pedestrian movements data measured with remote sensing technologies. In this regard, the endeavour to measure the pedestrian activity of a complex environment is a challenging task. Within an Urban Informatics approach (Foth et al. 2011) the acquisition of time-variant geolocated data is essential to describe and analyse the built environment, and its applications are in continuity with other research works already presented by the authors (Gorrini et al. 2021). In this case, the selection of the Square was not random: a web-accessible webcam is available on the top of one of the buildings facing the central part of the square. In particular, five different 30 minutes-long recordings were acquired on Thursday the 15th of July 2021. These footages were subsequently analysed by using the deep learning video analytics method YOLO v5 (Jocher et al. 2021) in order to detect the pedestrian locations in each frame, allowing for the comprehension of the utilization patterns of the space.

Furthermore, Piazza Duomo can be considered as one of the most important squares of the city given its central location, the historical relevance and its 'shopping destination' nature. The analysis of this space is of particular interest since it is an environment used for many different reasons (e.g., commuting, shopping, tourism, etc.) and it is structured as an interrupted pedestrian public space with several generators and attractors points (e.g., subway accesses, shopping anchors, cultural venues, etc.).

The paper is structured as follow: the following chapter will propose a review of the Space Syntax theories' debate, highlighting the opportunities and the limitation both in general and in relation to the research's scope, followed by a review of movement data acquisition techniques, and a comprehensive explanation of the methodologies employed both to assess the VGA metrics and to measure the footfall of the square. Lastly the results are presented with a discussion of the outcomes in relation with the research questions.

2 LITERATURE REVIEW

Space Syntax theories have been tested since their introduction in the effort to correlate the syntactic measures with empirical evidence, such as the recorded flows of pedestrians or vehicles. The outcomes of these studies led to different, and seldom inconsistent results, which fuelled a debate lasting for several years: a recent research article thoroughly reports the key points of the topic (Ericson et al. 2020).

In the case of the Grid-based analysis (i.e., *VGA*), the metrics were tested in various setting, among which museums (Turner and Penn 2002) and urban environment (Turner 2003), reporting *"a range of correlation values ranging from weak (e.g., .142) […] to strong (e.g., .98)"* (Ericson et al. 2020, page 1480), the same metrics were also used to assess the quality of public square in comparison to other qualitative factors (Gümüş and Erdönmez 2021). A notable urban environment case study are the experiments made on the Barnsbury area in London (Turner 2003), based on 116 gate counted pedestrian movements (Penn and Dalton 1994). In this case, the correlation coefficients ranged from .31 for the Visual Integration model to .67 for the Agent-based model and to .73 for the Multivariate agents/axial model. Moreover, in the case of agent simulation (Turner 2007), again on the Barnsbury dataset, the agents' correlation varies between .46 (3-step 170° field of view) to .62 (360° field of view), is it worth mentioning that the behaviour of the latter is consistent with the *Through Vision* metric.

Furthermore, methodological issues were also raised: it has been noted how the sampling grid resolution influences the outcomes of the VGA (Desyllas and Duxbury 2001) and, to the best of authors knowledge, there is an absence of literature about the role of *Restrict visible distance* parameter available in Depthmap. Moreover, as noted by (Koutsolampros et al. 2021), VGA metrics are hardly applicable to complex cases in a consistent manner, that is why it is proposed the utilization of a combination of singular measures in a multi-variate model (Koutsolampros et al. 2021). This approach is also explored in another research (Hacar et al. 2020), where a regression analysis is used to test various Line-Based and Grid-Based measures in respect to the pedestrian counted at gates of a university campus.

Finally, considering the gathering of empirical pedestrian data used in the previous research works, it is possible to distinguish two main categories: automatic and manual systems (Conroy 2001). Both can rely on sensing technologies to quantify space utilization, as outlined in *Table 1* (van Nes et al. 2021). Among them, videos represent the most versatile mean of gathering, as they can be mounted at a fixed location (i.e., snapshots, general movement traces, gate counts, mean occupancy), or mobile (i.e., pedestrian following). Wi-Fi and Bluetooth technologies can be used for mean occupancy analyses, while GPS/LBS sensors can be used to perform pedestrian following.

Table 1: The table show the classical data gathering methods and the available data outputs.

	Snapshots	*General movement traces*	*Gate counts*	*Walking with videos*	*Mean occupancy*	*Pedestrian following*
Images	•					
Videos	•	•	•	•	•	
Wi-Fi/ Bluetooth					•	
GPS/LBS						•

In addition to this, *Table 2* classifies the technologies presented in *Table 1A* based on automatic data counting methodologies, purposes and limitations and Space Syntax analyses typologies. Data counting methodologies can be divided in two categories,

(i) algorithm-based counting methods, namely those that rely on additional analyses to retrieve data. Images and videos fall under this category; here, automatic counting and analytical techniques based on deep learning algorithms enable the acquisition of detailed and continuous information at a large scale, overcoming the limitations linked to manual counting.

(ii) sensor-based counting methods, which depend on the technology embedded in the sensor to retrieve data. Wi-Fi, Bluetooth, GPS/ LBS technologies fall in this category.

The analysis purposes categorization, see *Table 2*, is structured in three categories, enabling the study of:

(i) spatial patterns, pedestrian behaviours and movements in an area (e.g., grouping, sitting, etc.) Spatial patterns can be gathered through images or videos by analysing spatial distribution and behaviours in a determined area), or through GPS/ LBS by means of repeated recognition of users at a given location and their spatial distribution.

(ii) densities, quantification of the footfall around hotspots. All sensors listed in *Table 2* can be used to retrieve densities around hotspots., (Soundararaj et al. 2020) used Wi-Fi probe requests to identify footfall patterns in retails, (Lin and Hsu 2014) proposed a review of methods and algorithms to mine mobility patterns from GPS data.

(iii) routes, the analysis of routing preferences (e.g., most walked paths, etc.) Routes can be retrieved with GPS/ LBS technologies, these allow for complete studies of routes chosen by users, or through videos, recognizing people's trajectories through deep learning based tracking analyses.

Limitations of sensors can be classified as:

(i) extension, i.e.,the analysis is restricted to a limited area, as for images and videos

(ii) users,i.e., the analysis is restricted to a limited number of people. as for Wi-Fi, Bluetooth, GPS technologies.

Lastly, the sensors are classified based on which Space Syntax analyses can be linked to them:

(i) line-based analyses relate to route preferences and traces, (e.g.,., videos, GPS/LBS)

(ii) grid-based analyses mainly relate to densities and patterns, (e.g., images, videos, Wi-Fi, Bluetooth)

A close application to the one proposed in this research (namely, the application of deep learning pedestrian recognition to the footage of a public space) is an occupancy study in an educational building (Tomé et al. 2015). In this case, video analytics is used to identify moving and standing individuals, to identify the utilizations patterns of the space. Furthermore, the research proposes a 'Difference Index' as a *"representation of the difference between the potential of movement previewed (VGA) and the observed movement"* (Tomé et al. 2015, page 93).

Table 2: The table shows data gathering techniques and automatic counting methods.

	Counting method		***Purpose***			***Limitations***			***Analysis type***	
	Sensor	Algorithm	Spatial patterns	Densities	Routes	Extension	Granularity	User sampling	Line based	Grid based
Images		•	•	•		•				•
Videos		•	•	•	•	•			•	•
Wi-Fi/Bluetooth	•			•			•	•		•
GPS/LBS	•		•	•	•			•	•	

3 METHODOLOGY

The following chapter describe the methodology used to (*i*) assess the VGA metrics; (*ii*) collect the pedestrian data with video analytics techniques for the case study.

3.1 Case study definition and VGA metrics selection

The analysis of Piazza Duomo in Milan (Italy) is of peculiar interest, since it is structured as an interrupted pedestrian space. The inner area of the square is interrupted both visually and physically by the subway accesses, by monumental lampposts and by the central equestrian statue. Apart from the geometry of the space itself, the surrounding area was also considered, up to a buffer of 250 meters *(see Figure 1)* from the central point of the analysed area. On the east-west direction several important shopping street merge in the Square, acting as visual corridors, and in north-south direction the gallery and the museal complex of Arengario create an uninterrupted linear system.

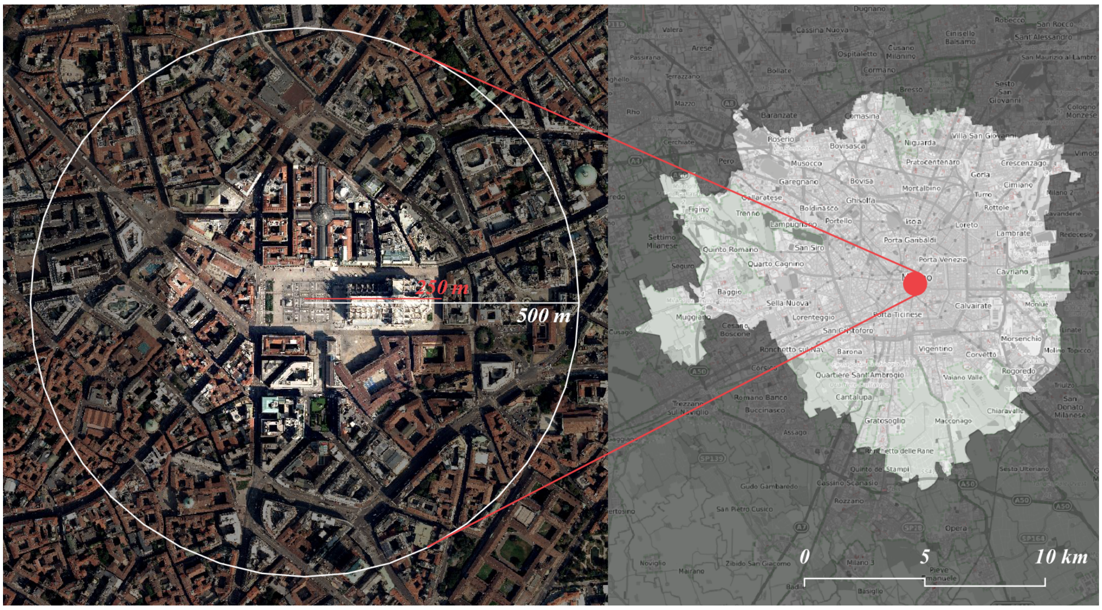

Figure 1: Location of case study and area considered for the preparation of the drawing of the pedestrian area.

The drawing used in Depthmap[1] is obtained from a collection of open geodata[2], which have been enriched to include pedestrian spaces not visible from satellite, but present in the analysis settings (e.g., galleries, covered alleys, etc.), and to introduce temporary obstacle present at the moment of the pedestrian flow monitoring. *Figure 2* shows the extent of the pedestrian are considered for the analysis.

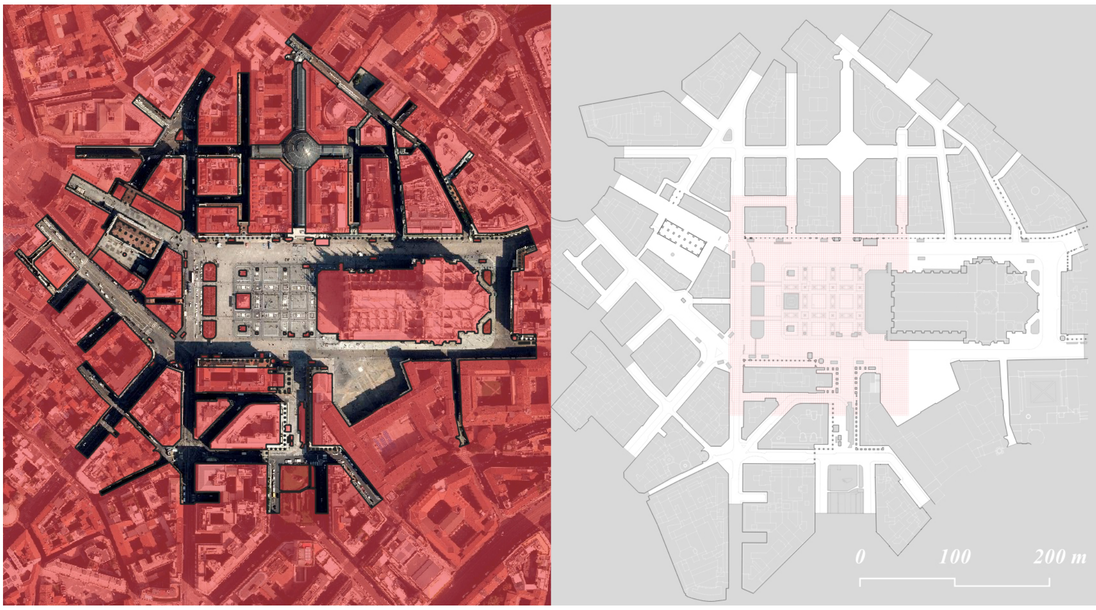

Figure 2: Extent of the area considered for VGA (left) and drawing used on Depthmap (right), with the 2x2m grid superimposed on the middle part of the square.

The analysed space is configured as an open-geometry area, that is expected to be used in diverse manners, as follows: (*i*) users crossing the central area to get to the other side of the square; (*ii*) users walking slowly and stopping in the middle part to appreciate the architecture of the square; and (*iii*) users walking in relation to urban functions (e.g., subway accesses, shopping venues, etc.). For these reasons, the authors selected several VGA metrics, listed in *Table 3*. In addition to this, the metrics are calculated with diverse restricted visibility distances (i.e., *NR – Not Restricted*, 150m, 100m and 50m), to test the outcomes against the measure pedestrian data.

Table 3: Selected VGA metrics, with respective properties and short description, for an extensive explanation see Koutsolampros et al. 2021.

	Metric	*Measuring*	*Extent*	*Short description*
Cn	Connectivity	Size	Local	Number of cells visible from a specific cell
I_Cm	Isovist Compactness	Shape	Local	Complexity of the isovist's shape
I_Dm	Isovist Drift Magnitude	Potential to explore	Local	Distance from a subject point to the centre of gravity of its isovist
I_O	Isovist Occlusivity	Potentiality to explore	Local	Proportion of unseen space that may be revealed during movement
TV	Through Vision	Potential to move	Local	Number of lines of visibility that pass through a specific cell
V_CC	Visual Clustering Coefficient	Potential to move	Local	Ratio of the number of cells in an isovist that can see each other to the total possible connections that could exist between those cells
V_Co	Visual Control	Control	Semi-global	Amount of space visible from a cell in relation to other directly visible cells
V_Cl	Visual Controllability	Control	Semi-global	Ratio between the number of visible cells (immediate neighbours) and the sum of all the cells visible from the immediate neighbours
V_In	Visual Integration (HH)	Normalised depth	Global	Normalised version of Mean Visual Depth
V_RE	Visual Relativized Entropy	Complexity	Global	Normalised measure of complexity of the overall system

3.2 Video description and analytics techniques

In the present study, five videos were analysed to obtain pedestrian footfall data. The footage was recorded on July 15, 2021, in five different moments of the day, specifically at 08:00, 11:00, 12:45, 15:00 and 18:00, each of them depicted a 30 minute time interval, with a size of 1920x1080 pixels and a frame rate of 15 frame per second (FPS).

Yolov5 (Jocher et al. 2021) with deep sort integration was used to detect pedestrians in the square, as visible in *Figure 3*. An open-source model was used, which was trained on webcam images in Montreal, Canada, obtaining a mean Average Precision (mAP) equal to 0.809 in pedestrians' recognition on the original training set[3]. The output is a text file with information on the frame number and x y pixel coordinates representing the location of the detected pedestrian. Specifically, the coordinate given to each detection is the middle point of the lower side of the bounding box, representing the location at ground of the pedestrian.

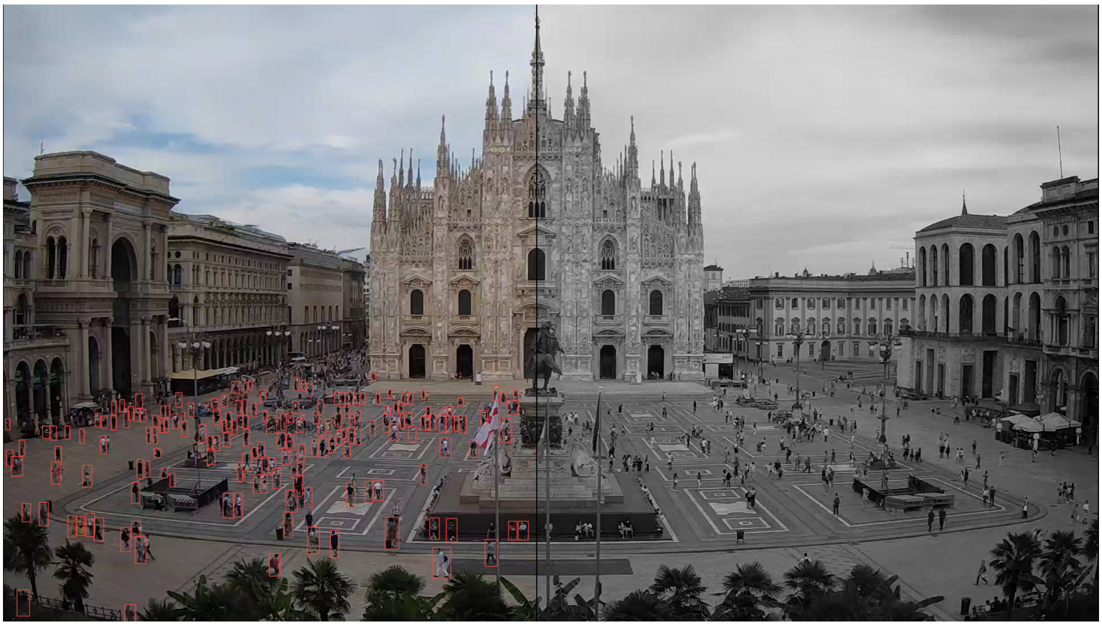

Figure 3 – Sample frame of the used footage, without detection (right) and with detected pedestrians (left).

3.3 Video footage georeferencing

In this phase, a georeferencing technique was implemented, in order to estimate the geographical position of the distribution of pedestrians in the square. The methodology procedure is outlined in the following points:

1. One frame of the video was imported in a GIS software, specifically QGIS[4], and georeferenced using the Thin Plate Spline (TPS) as transformation type as algorithm. Given the distortion of the footage, roughly 240 Ground Control Points (GCPs) were selected to anchor points in the perspective image.
2. A unique random colour image was created, with the same size as the original image (1920x1080 pixels). Here, each random RGB value acted as pixel identifier (~2,000,000 unique pixels), allowing for a one-to-one match between the two images.
3. The previously exported GCPs were used to georeferenced the random RGB pixels image, using the Nearest Neighbour resampling method to maintain original RGB values without resampling-generated artifacts.
4. Lastly, the pixels of the original and georeferenced images were matched using a python algorithm written by the authors. The goal was to maintain the mathematical complexity obtained through the georeferencing algorithm in the transformation of all the detections.

In order to assess the deformation obtained through the georeferentiation process, the spatial accuracy was quantified through the size of the pixels in the georeferenced image (0.28x0.28m). Then, two metrics were developed by matching RGB values in the two images. Here three cases are identified:

1 One-to-one pixels – these are pixels that present once in the perspective and in the georeferenced images;

2 Compressed pixels – these are pixels that are present in the perspective image that do not have a correspondence in the georeferenced one, shown in *Figure 4*;

3 Stretched pixels – these are pixels in the perspective image that present more than one correspondence in the georeferenced image, shown in *Figure 5*.

To locate stretched pixels, the uniqueness of RGB values in the georeferenced image was checked. Between the original and the georeferenced frames, a one-to-one correspondence is present for pixels in the left part of *Figure 4*, while deformation increases in areas where GCPs could not be placed (namely, behind the equestrian statue).

To locate compressed pixels, first one-to-one pixels' correspondences were identified. Thus, a nearest neighbours' method was used to match not georeferenced pixels to the nearest georeferenced one in the original image. This step allows for the complete transformation of the original image, enabling positioning of all detections. *Figure 5* depicts the number of pixels that are assigned to their nearest georeferenced neighbour.

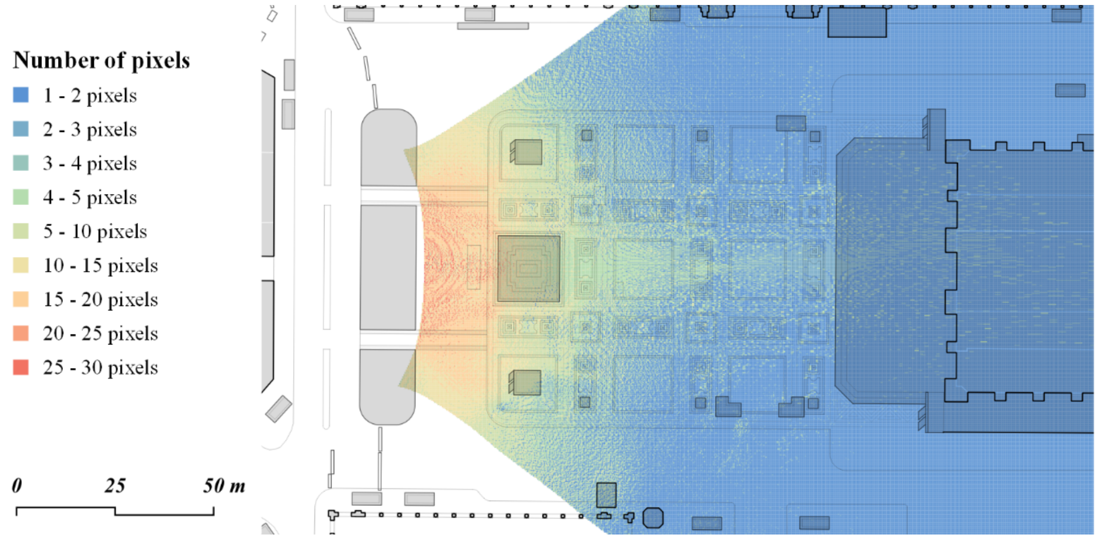

Figure 4 – The image shows the compressed pixels, namely the pixels that represent more than one pixel of the original image

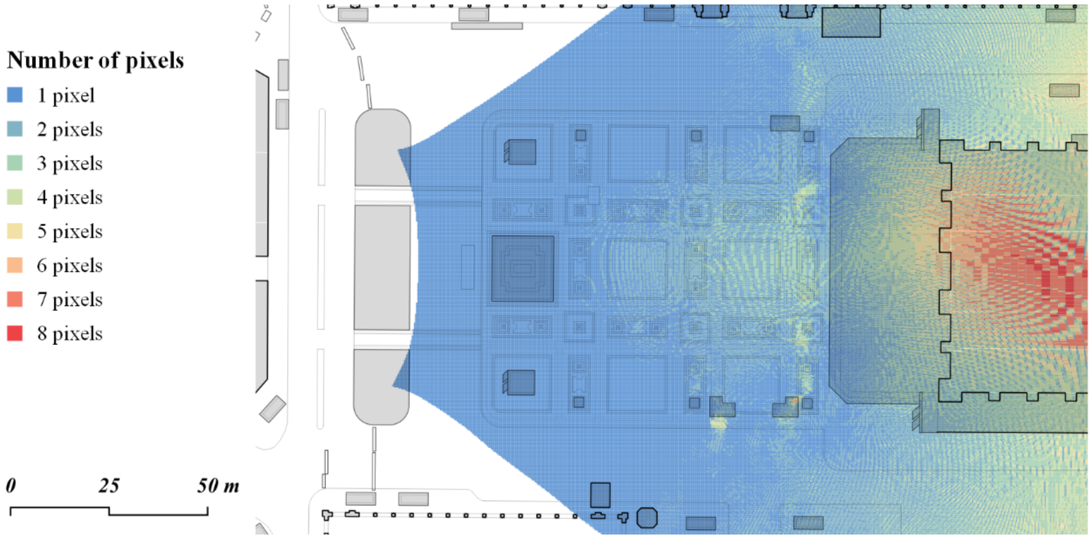

Figure 5 – The image shows the stretched pixels, namely the pixels that are represented by more than one pixel of the original image. This metric can be interpreted as a measure of accuracy of the georeferencing process, highlighting how many pixels are considered as one in the transformed image.

3.4 Detections' measurements

After the analysis of the videos to identify the pedestrian and after the development of the georeferencing method for the footage's frames, the footfall quantification was outlined as follows:

1. The georeferenced pedestrian coordinates (x, y) in each frame were associated to the corresponding cell of the VGA grid, visible in *Figure 2*.
2. A mean occupancy value per second was computed, averaging the number of detections that fell into the same grid cell within fifteen frames.
3. Third, a cumulated footfall measure was computed. The mean occupancy $(\overline{Oc}_{30})$, expressed as *people/sec* was summed progressively in thirty seconds bins, obtaining increasingly accurate aggregates starting from 0-30 seconds to 0-1,800 seconds.
4. Finally, another metrics was introduced, with the goal to identify the cells with a higher turnover with respect to the cells used to stay still or wait. This measure of variation (Var) is derived from the mean occupancy computed on five seconds bins $(\overline{Oc}_5)$: each bin is compared with the previous to measure the deltas between the two in terms of *people/sec* $(\Delta\overline{Oc}_5)$ and in terms of dummy variable (Ga) (i.e., the cell was activated). The sum of the deltas and the count of the dummy variables are summed and divided by the mean occupancy of each cell, as expressed in *Equation 1*.

$$Var = \left(\sum_{n=1}^{n} \Delta\overline{Oc}_i + \sum_{n=1}^{n} Ga_i\right)\frac{1}{Oc_n}$$

Equation 1 – The equation expresses the formula to compute the turnover of each grid cell for the 5 seconds bins, normalizing the result on the cell's final mean occupancy, at the bin *n = 360*.

4 RESULTS

The following chapter is organized as follows: firstly, an outline of the results of the video analytics process is presented, describing the measured pedestrian footfall and movement metric; secondly, the outcomes obtained with the VGA metrics are described; finally, the correlation analyses between the two former datasets are shown.

4.1 Measured pedestrian footfall

The video analytics process described before allowed for the calculation of the occupancy of the space on the 2x2m grid. Since the footage was taken at different hours of the day, the footfall measured in the videos is varying in magnitude and patterns, reflecting the varying nature of the public space. *Table 4* shows the results of the video analytics, including descriptive statistics to highlight the diversity of the results.

Table 4 – Summary of the measured pedestrian footfall. The number of cells considered in the video is n = 2.060, meanwhile the number of cells included in the Footfall Data (FD) and Movement Data (MD) datasets is n = 1,921.

Time	*Min*	*Max*	*Mean*	*St. Dev.*	*1st perc.*	*99th perc.*
08:00	1.133	1,126.800	45.240	49.888	3.267	194.301
11:00	0.800	1,422.467	181.575	122.239	14.439	620.856
12:45	2.933	1,491.400	212.804	132.265	24.318	704.754
15:00	3.400	1,527.933	199.757	129.567	26.457	655.737
18:00	3.200	1,504.333	211.024	132.102	24.663	704.367
FD	144.000	2,499.667	824.395	352.383	-	-

In addition to this, it is possible to estimate the number of pedestrians detected by considering the average number of detections per frame of each video: (08:00 – 88.925 pedestrians, 11:00 – 305.677 pedestrians; 12:45 – 370.363 pedestrians, 15:00 – 344.902 pedestrians, and 18:00 – 352.594 pedestrians).

Based on these datasets, the authors elected to remove the outliers from the distributions, considering the 1st and the 99th percentile of each video sequence, then, two metrics were defined: *(i)* Footfall Data (FD) and *(ii)* Movement Data (MD). The former, defined in *Equation 2*, is the sum of the cumulated mean occupancy for the five videos.

$$FD_t = Oc_{t\,(08:00)} + Oc_{t\,(11:00)} + Oc_{t\,(12:45)} + Oc_{t\,(15:00)} + Oc_{t\,(18:00)}$$

Equation 2 – Footfall Data (FD) is calculated as the sum of the cumulated occupancies *(t = 1,800)* of the five videos, calculated for each grid cell (n = 1,921)

The latter, defined in *Equation 3*, is the mean value of the measure of variation (*Var*) for the five videos.

$$MD = \frac{Var_{(08:00)} + Var_{(11:00)} + Var_{(12:45)} + Var_{(15:00)} + Var_{(18:00)}}{5}$$

Equation 3 – Movement Data (MD) is calculated as the mean value of the variation metric for the five videos, calculated for each grid cell (n = 1,921).

Footfall Data (FD), shown in *Figure 6* and *Table 3*, represents the utilization pattern of the square, and it was used since it describes an average situation of different moments of the day, diminishing the influence of functions and generators / attractors points. Movement Data (MD), shown in *Figure 7*, represents the cells with a high turnover, namely the areas where the pedestrians don't stop but generally keep moving.

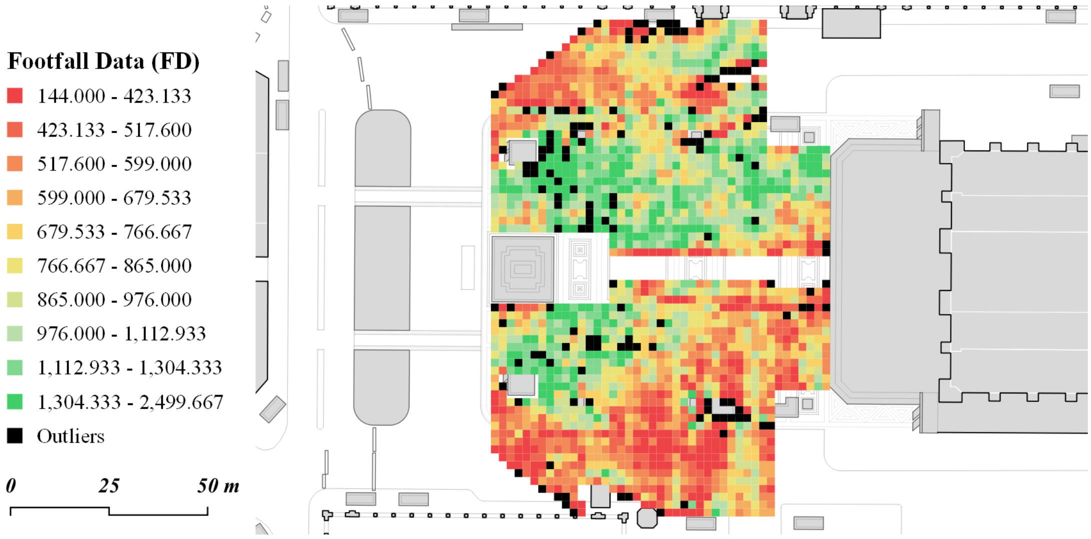

Figure 6 – Map showing the Footfall Data (FD) on the grid, with the outliers highlighted in black.

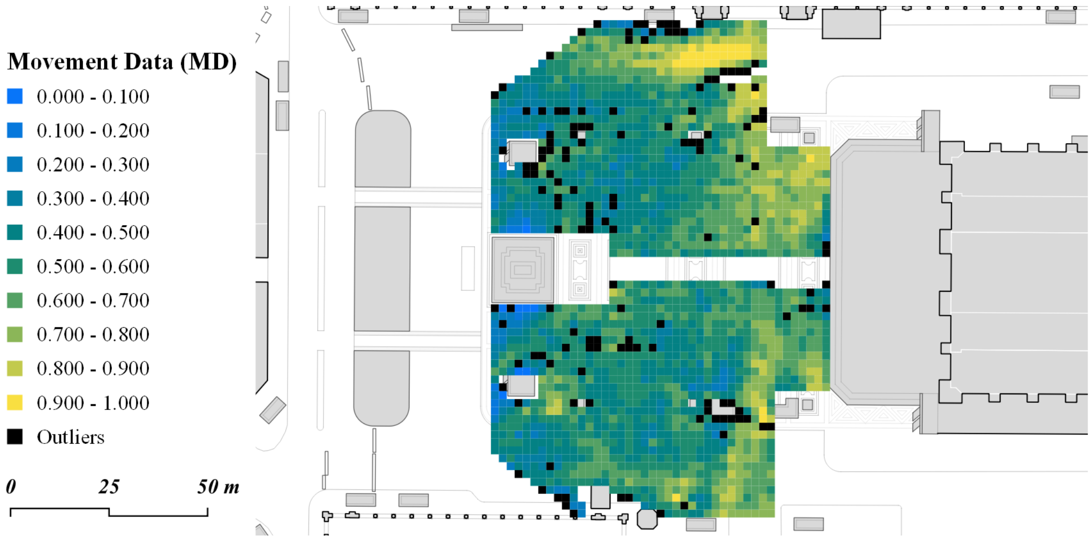

Figure 7 – Map showing the Movement Data (MD) on the grid, with the outliers highlighted in black.

4.2 VGA metrics

The VGA analyses were calculated for the area shown in *Figure 2*, for several maximum visibility distances (i.e., *NR – Not Restricted*, 150m, 100m and 50m), in order to compare the results in the effort to understand which distance can describe effectively an open public space. In *Figure 8*, it is possible to see how out of the ten selected metrics, three of them (i.e., I_Cm – Isovist Compactness, I_DM – Isovist Drift Magnitude, and I_O – Isovist Occlusivity) are not related to the maximum visibility distance parameters, since those are based on the un-restricted isovist's shape. The other metrics show a consistent difference among the four distances, namely the transition from the large-scale topology to the open public space topology.

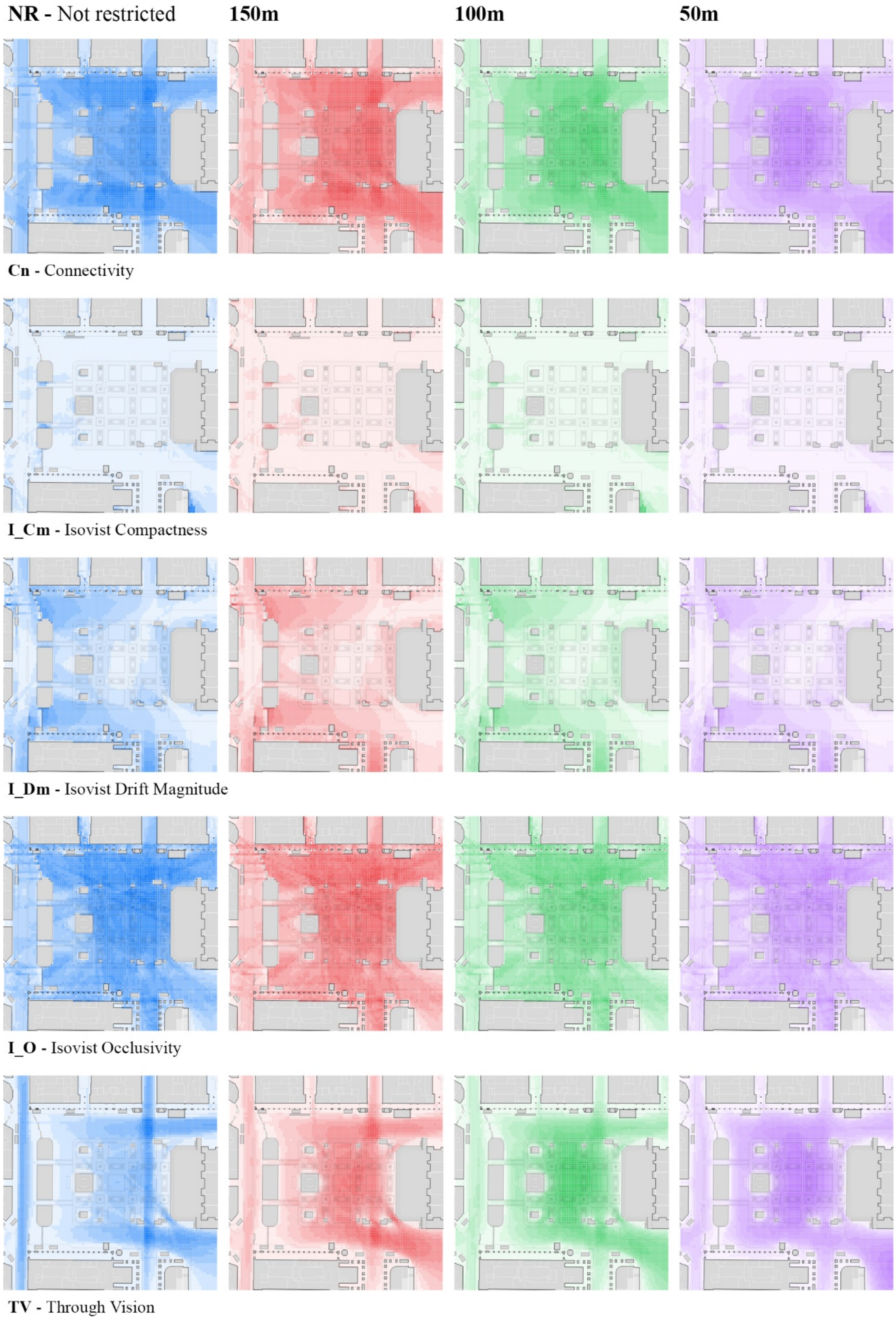

Figure 8 – VGA metrics with the variation of the restricted visibility parameters. All the values are shown as regular intervals from the minimum (lighter colour) to the maximum value stronger colour). *Figure 8* continues in the next page.

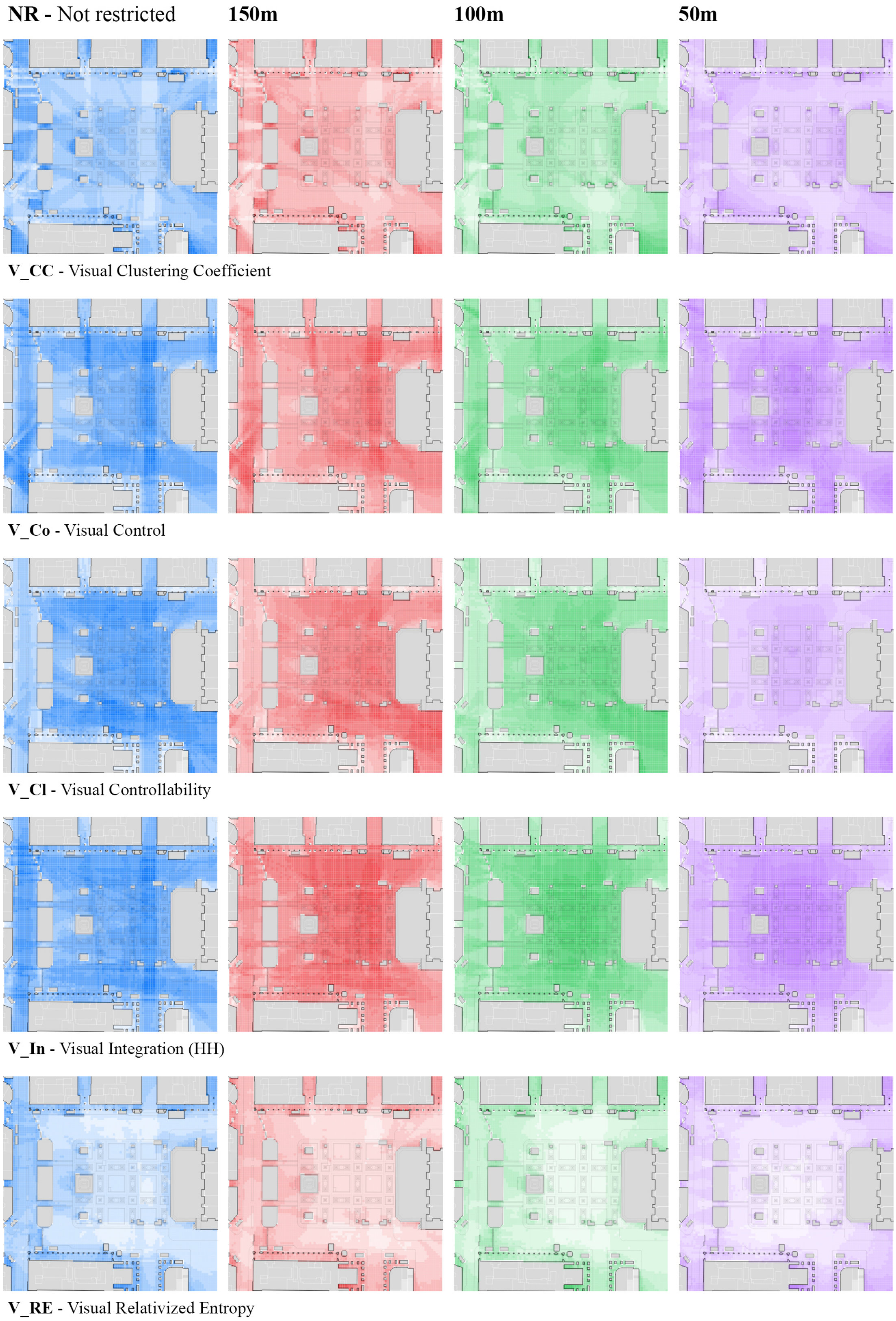

Figure 8 – VGA metrics with the variation of the restricted visibility parameters. All the values are shown as regular intervals from the minimum (lighter colour) to the maximum value (stronger colour). *Figure 8* starts in the previous page.

4.3 Correlation analyses

Finally, the VGA metrics, with the variation of the restricted visibility distance parameter, and the Footfall Data (FD) were normalized on a 0 – 1 scale, creating Z-Scores that follow the normal distribution of the values. This normalization was done in order to lower the weight of the higher and lower values for both the metrics and the pedestrian footfall. On the contrary, the Movement data (MD) was linearized on a 0 – 1 scale, to facilitate the readability of the visualization. Then, using Pearson's Correlation Coefficient, the datasets were analysed with a series of correlations analyses[5]. The results for the Footfall Data (FD) and the Movement Data (MD) are shown, respectively in *Table 5* and *Figure 9*, and in *Table 6* and *Figure 10*.

Table 5: Footfall Data (Z_FD) correlation analysis results. Not-significant p-values are highlighted in bold. The number of samples for all correlations is n = 1921.

Metric		NR		150m		100m		50m	
		r	*p < .001*	*r*	*p < .001*	*r*	*p < .001*	*r*	*p < .001*
Z_Cn	Connectivity	*-.265*	•	*-.085*	•	*.155*	•	*.287*	•
Z_I_Cm	Isovist Compactness	*-.234*	•	*-.234*	•	*-.234*	•	*-.234*	•
Z_I_Dm	Isovist Drift Magnitude	*-.376*	•	*-.376*	•	*-.376*	•	*-.376*	•
Z_I_O	Isovist Occlusivity	*.176*	•	*.176*	•	*.176*	•	*.176*	•
Z_TV	Through Vision	*-.054*	•	*.098*	•	*.211*	•	*.174*	•
Z_V_CC	Visual Clustering Coefficient	*.259*	•	*.228*	•	*.154*	•	*-.117*	•
Z_V_Co	Visual Control	*-.226*	•	*-.081*	•	*.122*	•	*.117*	•
Z_V_Cl	Visual Controllability	*-.268*	•	*-.196*	•	*.037*	***.185***	*.280*	•
Z_V_In	Visual Integration (HH)	*-.155*	•	*.089*	•	*.262*	•	*.326*	•
Z_V_RE	Visual Relativized Entropy	*.306*	•	*.075*	•	*-.133*	•	*-.048*	•

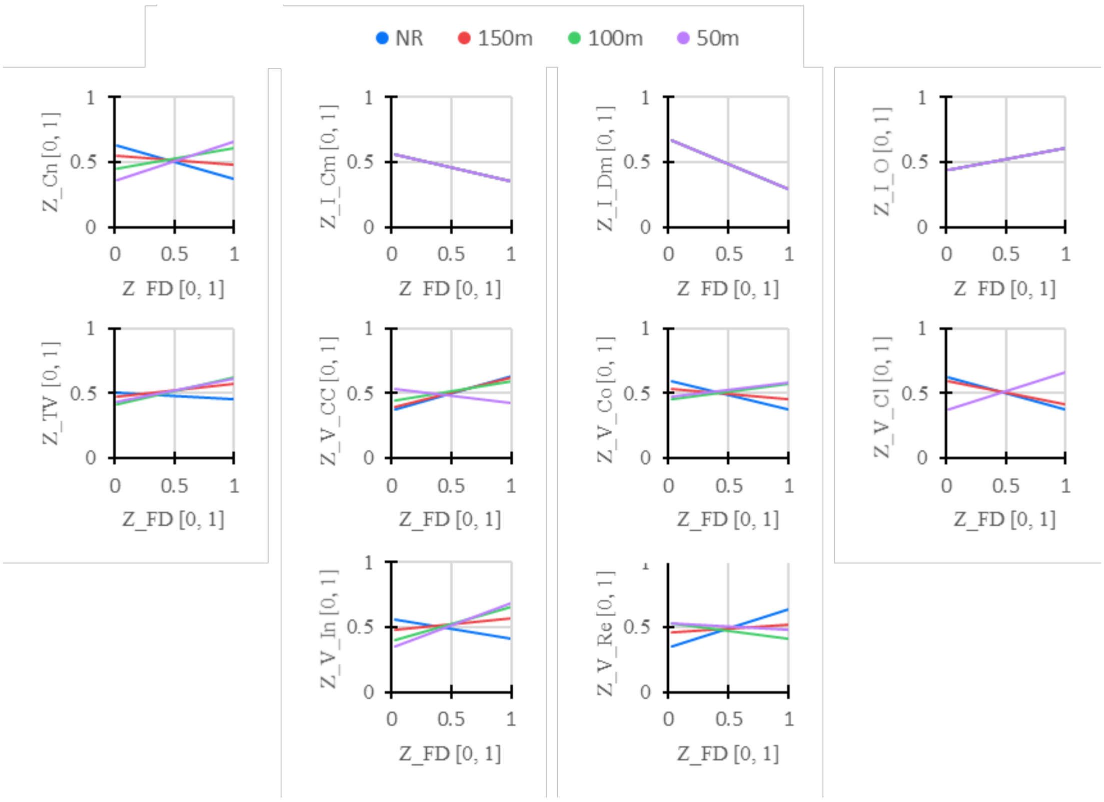

Figure 9 – Linear trend lines showing (when significant) the correlation results between the normalized Footfall Data (Z_FD) and the normalized values of the VGA metrics.

Table 6: Movement Data (MD) correlation analysis results. Not-significant p-values are highlighted in bold. The number of samples for all correlations is n = 1921.

Metric		***NR***		***150m***		***100m***		***50m***	
		r	***p < .001***	***r***	***p < .001***	***r***	***p < .001***	***r***	***p < .001***
Z_Cn	Connectivity	*.373*	•	*.411*	•	*.277*	•	*-.062*	*.007*
Z_I_Cm	Isovist Compactness	*-.244*	•	*-.244*	•	*-.244*	•	*-.244*	•
Z_I_Dm	Isovist Drift Magnitude	*-.122*	•	*-.122*	•	*-.122*	•	*-.122*	•
Z_I_O	Isovist Occlusivity	*.265*	•	*.265*	•	*.265*	•	*.265*	•
Z_TV	Through Vision	*.373*	•	*.156*	•	*-.006*	***.891***	*-.086*	•
Z_V_CC	Visual Clustering Coefficient	*-.326*	•	*-.391*	•	*-.295*	•	*.124*	•
Z_V_Co	Visual Control	*.402*	•	*.468*	•	*.327*	•	*-.120*	•
Z_V_Cl	Visual Controllability	*.297*	•	*.394*	•	*.383*	•	*-.032*	*.166*
Z_V_In	Visual Integration (HH)	*.396*	•	*.303*	•	*.091*	•	*-.114*	•
Z_V_RE	Visual Relativized Entropy	*-.281*	•	*-.433*	•	*-.145*	•	*.210*	•

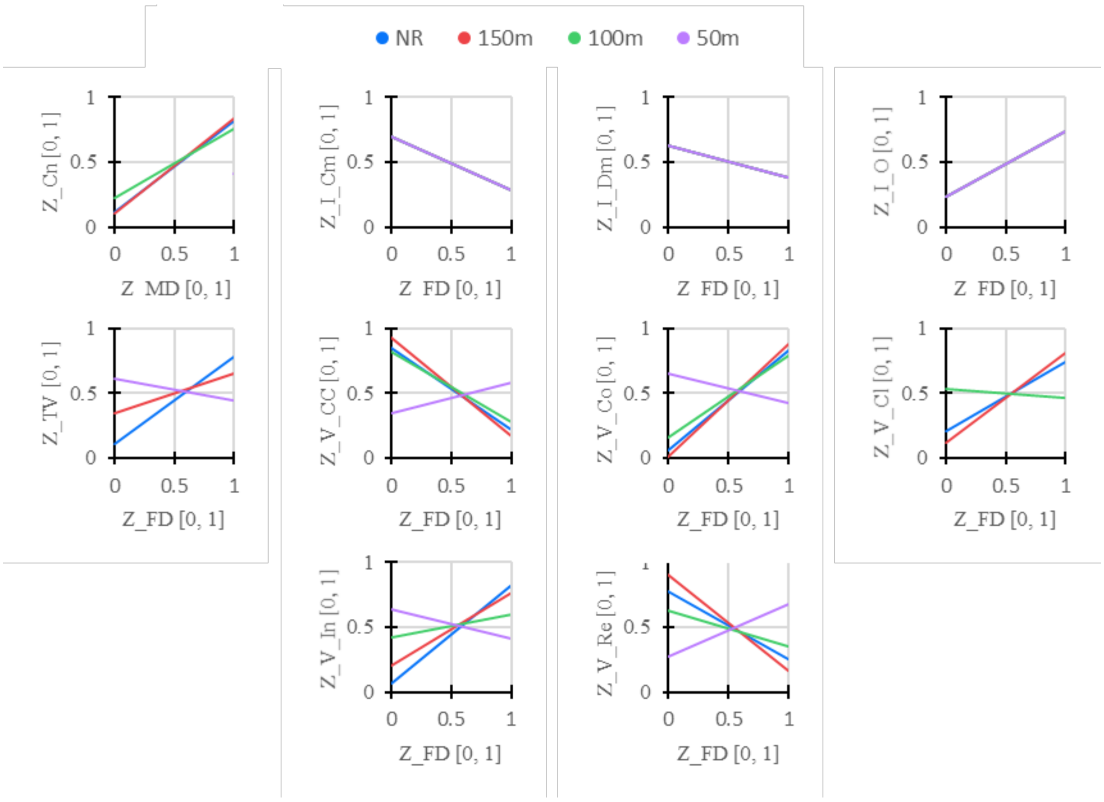

Figure 10 – Linear trend lines showing (when significant) the correlation results between the normalized Movement Data (MD) and the normalized values of the VGA metrics.

5 DISCUSSION

The utilization of video analytics analyses on a webcam located in a large public space allowed for the measure the footfall in five different moments of the day. This dataset was used to define two metrics describing the space utilization (i.e., Footfall Data (FD) and Movement Data (MD)), that were tested against a selection of VGA metrics, in order to assess the presence of a relation between the former and the latter data, using a correlation analysis methodology.

The detections, and thus the pedestrian utilization of the square, are different during the day. From a qualitative preliminary analysis, the morning (08:00) is the most different, because the number of pedestrians present in the square is very limited and their movements follow desire lines connecting different urban functions. Meanwhile, the other datasets are more consistent with each other, showing both desire line shaped patterns and a “natural” utilization of the open space, using mostly the central part of the square.

To confirm this hypothesis, a series of independent samples t-tests (two-tailed) was performed. This analysis showed a significant difference between the average number of pedestrians detected per time intervals: (*i*) 08:00 and 11:00 ($p < .01$); (*ii*) 08:00 and 12:45 ($p < .01$); (*iii*) 08:00 and 15:00 ($p < .01$); (*iv*) 08:00 and 18:00 ($p < .01$); (*v*) 11:00 and 12:45 ($p < .01$); (*vi*) 11:00 and 15:00 ($p < .01$); (*vii*) 11:00 and 18:00 ($p < .01$); (*viii*) 12:45 and 15:00 ($p < .01$); (*ix*) 15:00 and

18:00 ($p < .01$). Results showed also a non-significant difference between the time intervals 12:45 and 18:00 ($p > .05$). This is based on the observed similar values during peak hours in the morning and the afternoon *(see Table 3)*.

The correlation results for the Footfall Data (FD), visible in *Table 5* and *Figure 9,* do not show strong correlations between the analysed data. The values range from negative to positive correlations, but the absolute values are never higher than .376, in the case of the *Isovist Drift Magnitude*. One of the research questions was to understand the impact of the variation of the restricted visibility parameter, which show significant outcomes in the correlation analysis. In *Figure 7* is possible to see how shorter radii (i.e., 100m and 50m) result in higher values in the centre of the public space, specifically for the following metrics: (*i*) *Connectivity*, (*ii*) *Through Vision*, (*iii*) *Visual Control*, and (*iv*) *Visual Integration (HH)*.

Moreover, the correlation analysis results for the Movement Data (MD), shown in *Table 6* and *Figure 10* express different outcomes. The correlation coefficient are still moderate at best, such as the cases of *Connectivity* (.373 – NR and .411 – 150m), *Through Vision* (.373 – NR), *Visual Clustering Coefficient* (-.326 – NR and -.391 – 150m), *Visual Control* (.402 – NR and .468 – 150m), and *Visual Integration (HH)* (.396 – NR and .303 – 150m), but these results are more in line with the expected results with, as example, *Visual Integration* identifying space linked to movements. With respect to this dataset, the variation of the restricted visibility parameter demonstrates how larger radii (i.e., NR and 150m) lead to different results, with higher correlation values with the VGA metrics.

The authors' interpretation of these results is that in this case study the footfall patterns are strongly linked to a case-specific utilization of the space. The touristic nature of the square and the presence of the subway accesses lead to a higher cumulative density in the middle part of the space, but, at the same time, the area utilized to move around draw some "desire lines", especially near the gallery, that are also well represented in the VGA analyses. The Movement Data (MD) metric allow to clean the recorded movement data from the noise of people standing still in the space, narrowing the analysis to the actual flows.

In support of this statement, it is possible to see how the short-radius *Visual Integration* (50m), which shows a positive correlation of .326 with the Footfall Data (FD), is not linked to the topological shape of the larger environment, but it is very similar to results obtainable in a perfect square shape: higher values in the middle, with lower values toward the edges, also the *Connectivity* follow a very similar pattern. Meanwhile, the Movement Data (MD) is better correlated with the large-radius and, thus, showing how the environment actually shapes the presence of pedestrian flows. Finally, in the case of *Through Vision*, which can be regarded as an alternative to EVA agents for 360° field of view (Turner 2007), the comparison between the

Footfall Data (FD) and Movement Data (MD) shows much more consistent results in the latter case, also considering the interpretation of the metric (Turner and Penn 2002).

The outcomes of these analyses are in line with other recent research studies (Ericson et al. 2020, Koutsolampros et al. 2021), and show limitations in the application of VGA metrics in relation to different typologies of spaces. However, the proposed case study is a particularly complex environment chosen to stress the reliability of the VGA metrics in a space characterized by an inherent asymmetry of functions (the northern part is a continuation of one of the major shopping streets in the city) and the presence of strong attractors and generators, namely the subway accesses. These factors weight heavily in the pedestrian movements, greatly influencing the cumulated density as expressed in the FD dataset. On the other hand, the utilization of a "movement-based" metric, expressed in the MD dataset, allow to partially re-focus the analysis on the movements, showing the patterns not related to the staying still.

6 CONCLUSIONS AND FUTURE STUDIES

The research aims to contribute to the debate regarding Space Syntax metrics, in particular the Grid-based analysis VGA. The proposed case study is utilized to test the VGA metrics values, obtained by varying the restrict visibility parameter, in the central part of Piazza Duomo, Milan (Italy). The central portion of the public space was recorded through a webcam, and the footage was studied through video analytics techniques, measuring the pedestrian activity of the space discretized in a 2x2m grid.

The metrics and the pedestrian movements were tested in correlation analyses, showing weak to moderate correlation values. As thoroughly explained in the previous chapter, the square is a complex environment and should be regarded as a "stress test" for the proposed analyses. In this framework, the proposed Movement Data (MD) shows positive moderate correlation with VGA metrics usually associated to movements, namely *Connectivity, Through Vision* and *Visual Integration (HH).* As future improvement, this dataset can be refined by introducing tracking technologies in the video analytics techniques, associating a unique ID to each pedestrian and recording its location and movement speed. Introducing the speed as a factor, it is possible to identify more precisely flowing and staying behaviour, as seen in (Tomè et al. 2015).

However, it is worth mentioning how from preliminary analyses, not included in this article, and from the scientific debate (Turner 2003, Koutsolampros et al. 2021), it appears how additional factors could weight in the significance of the VGA metrics. Specifically, the Agent Analysis (Turner 2003), based on the VGA metrics, considers a series of access gates, acting as starting points of space exploration, and it is regarded as a more reliable analysis in the case of points of attraction and generation. Another approach, yet to be thoroughly tested, is the implementation of anchors-based metrics, specifically a metric value representing the distance from the main anchors. The weighting of the VGA metric, especially the *Visual Integration (HH)* with the distance from selected points could lead to a more reliable representation of the public space, in the effort to build a calibrated VGA model.

Lastly, especially in the proposed case of Piazza Duomo, future work will be focused on the possibility to further characterized pedestrian profiles by using proxemics, speeds and trajectories data collected through tracking video analytics techniques. In particular, according to a previous work already presented by the authors (Gorrini et al. 2016), it is proposed a general conceptualisation of different walking behaviours of single pedestrians and groups which was applied to investigate the impact of visual perception during locomotion (Gibson 2014) and pedestrian proxemic behaviour (Hall 1966) on the actual use of the considered public space:

- Time driven pedestrians: people who have time constraints and walk through a certain environment constantly adjusting the trajectory between origin and destination to preserve a high speed (very often singles and commuters accessing public transport services);
- Space driven pedestrians: tourists or shoppers who visit for the first time a certain environment or have an exploratory attitude. Sometimes they are organized in large groups led by a guide. They stop more often, either for taking pictures of interesting spots or for shopping;
- Social driven pedestrians: strollers and inhabitants who amble thought a certain environment since they live or work nearby the area (more often small groups or families). They can stop their walk for an improvised conversation with somebody they know or for looking at the shop windows.

REFERENCES

Becker, R., C´aceres, R., Hanson, K., Isaacman, S., Loh, J.M., Martonosi, M., Rowland, J., Urbanek, S., Varshavsky, A. and Volinsky, C. (2013). Human mobility characterization from cellular network data. Communications of the ACM, 56(1): 74–82. doi: 10.1145/2398356.2398375

Benedikt, M. L. (1979). 'To take hold of space: isovists and isovist fields.' *Environment and Planning B: Planning and design*, *6*(1), 47-65. doi: 10.1068/b060047

Conroy, R. (2001). *Spatial navigation in immersive virtual environments*. University of London, University College London (United Kingdom).

Buch, N., Velastin, S.A. and Orwell, J. (2011). A review of computer vision techniques for the analysis of urban traffic. IEEE Transactions on Intelligent Transportation Systems, 12(3): 920–939. doi: 10.1109/TITS.2011.2119372

Dalton, R & Dalton, N (2001). 'OmniVista: an application for isovist field and path analysis.' in *Proceedings, 3rd International Space Syntax Symposium*. Georgia Institute of Technology Atlanta.

Desyllas, J., & Duxbury, E. (2001, May). 'Axial maps and visibility graph analysis.' In *Proceedings, 3rd International Space Syntax Symposium* (Vol. 27, pp. 21-13). Georgia Institute of Technology Atlanta.

Ericson, J. D., Chrastil, E. R., & Warren, W. H. (2020). 'Space syntax visibility graph analysis is not robust to changes in spatial and temporal resolution.' *Environment and Planning B: Urban Analytics and City Science*, *48*(6), 1478-1494. doi: 10.1177/2399808319897624

Foth, M., Choi, J. H. J., & Satchell, C. (2011). 'Urban informatics.' In *Proceedings of the ACM 2011 conference on Computer supported cooperative work* (pp. 1-8). doi: 10.1145/1958824.1958826

Gibson, J. J. (2014). *The ecological approach to visual perception: classic edition*. Psychology Press.

Gorrini, A., Messa, F., Ceccarelli, G., & Choubassi, R. (2021). 'Covid-19 pandemic and urban mobility in Milan. Wi-Fi sensors and location-based data.' *TeMA-Journal of Land Use, Mobility and Environment*, *14*(2), 211-226. doi: 10.6093/1970-9870/7886

Gorrini, A., Vizzari, S., Bandini, S. (2016). 'Age and Group Driven Pedestrian Behavior: from

Observations to Simulations.' *Collective Dynamics*, 1(3), 1-16. doi: 10.17815/CD.2016.3

Gümüş, İ., & Erdönmez, E. (2021). 'Impact of spatial configuration to spatial quality: Venice and Istanbul.' *Journal of Architecture and Urbanism*, *45*(2), 205-216. doi: 10.3846/jau.2021.14306

Hacar, Ö. Ö., Gülgen, F., & Bilgi, S. (2020). 'Evaluation of the Space Syntax Measures Affecting Pedestrian Density through Ordinal Logistic Regression Analysis.' *ISPRS International Journal of Geo-Information*, *9*(10), 589. doi: 10.3390/ijgi9100589

Hall, E. T. (1966). *The hidden dimension*. Garden City, NY: Doubleday.

Hillier, B. (1999). 'The hidden geometry of deformed grids: or, why space syntax works, when it looks as though it shouldn't.' *Environment and Planning B: planning and Design*, *26*(2), 169-191. doi: 10.1068/b4125

Hillier, B. (2007). *Space is the machine: a configurational theory of architecture*. Space Syntax.

Hillier, B., & Hanson, J. (1984). *The Social Logic of Space*. Cambridge, New York: Cambridge University Press.

Hillier, B., Major, M. D., Desyllas, J., Karimi, K., Campos, B., & Stonor, T. (1996). Tate Gallery, Millbank: A study of the existing layout and new masterplan proposal.

Jocher G., Stoken A., Chaurasia A., Borovec J., NanoCode012, TaoXie, Kwon Y., Michael K., Changyu L., Fang J., Abhiram V., Laughin, tkianai, xyNONG, Skalski P., Hogan A., Nadar J., imyhxy, Mammana L., … & wanghaoyang0106 (2021). ultralytics/yolov5: v6.0 - YOLOv5n 'Nano' models, Roboflow integration, TensorFlow export, OpenCV DNN support (v6.0). Zenodo. doi: 10.5281/zenodo.5563715

Koutsolampros, P., Sailer, K., Varoudis, T., & Haslem, R. (2019). Dissecting Visibility Graph Analysis: The metrics and their role in understanding workplace human behaviour. In *Proceedings of the 12th International Space Syntax Symposium* (Vol. 12). International Space Syntax Symposium.

Lin, M. and Hsu, W.J. (2014). Mining gps data for mobility patterns: A survey. Pervasive and mobile computing, 12:1–16. doi: 10.1016/j.pmcj.2013.06.005

Penn, A. & Dalton, N. (1994). *The Architecture of Society: Stochastic Simulation of Urban Movement, Simulating Societies: The Computer Simulation of Social Phenomena*, UCL Press, London, pp. 85–125.

Sapiezynski, P., Stopczynski, A., Gatej, R. and Lehmann, S. (2015). Tracking human mobility using wifi signals. PloS one, 10(7): e0130824. doi: 10.1371/journal.pone.0130824

Soundararaj, B., Cheshire, J. and Longley, P. (2020). Estimating real-time high-street footfall from wi-fi probe requests. International Journal of Geographical Information Science, 34(2): 325–343. https://doi.org/10.1080/13658816.2019.1587616

Tomé, A., Kuipers, M., Pinheiro, T., Nunes, M., & Heitor, T. (2015). 'Space–use analysis through computer vision.' *Automation in Construction*, *57*, 80-97. doi: 10.1016/j.autcon.2015.04.013

Turner, A. (2003). ‘Analysing the visual dynamics of spatial morphology.’ *Environment and Planning B: Planning and Design*, *30*(5), 657-676. doi: 10.1068/b12962

Turner, A. (2001). ‘Depthmap: a program to perform visibility graph analysis.’ In *Proceedings of the 3rd International Symposium on Space Syntax* (Vol. 31, pp. 31-12).

Turner, A., & Penn, A. (2002). ‘Encoding natural movement as an agent-based system: an investigation into human pedestrian behaviour in the built environment.’, *Environment and planning B: Planning and Design*, *29*(4), 473-490. doi: 10.1068/b12850

Turner A. (2007). ‘To move through space: lines of vision and movement’ in *Proceedings, 6th International Space Syntax Symposium,* 12 – 15 June 2007. 6th Internal Space Syntax Symposium. URL: http:// eprints.ucl.ac.uk/4111/ (visited on February 2022)

Turner, A., & Penn, A. (2002). ‘Encoding natural movement as an agent-based system: an investigation into human pedestrian behaviour in the built environment.’, *Environment and planning B: Planning and Design*, *29*(4), 473-490. doi: 10.1068/b12850

Turner, A., Doxa, M., O'sullivan, D., & Penn, A. (2001). ‘From isovists to visibility graphs: a methodology for the analysis of architectural space.’ *Environment and Planning B: Planning and design*, *28*(1), 103-121. doi: 10.1068/b2684

van Nes, A., & Yamu, C. (2021). Empirical Data Collection and Analysis, and Connecting Data with Space Syntax. In A. van Nes, & C. Yamu (Eds.), Introduction to Space Syntax in Urban Studies (pp. 133-170). Springer. doi: 10.1007/978-3-030-59140-3_5