Automation in Construction xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Automation in Construction

journal homepage: www.elsevier.com/locate/autcon

Automated classification of construction site hazard zones by crowd-sourced integrated density maps

Heng Li[a], Xincong Yang[a,b,*], Martin Skitmore[c], Fenglai Wang[d], Perry Forsythe[e]

[a] Dept. of Building and Real Estate, The Hong Kong Polytechnic Univ., Hong Kong
[b] School of Civil Engineering, Harbin Institute of Technology, China
[c] School of Civil Engineering and the Built Environment, Queensland University of Technology (QUT), Australia
[d] School of Civil Engineering, Harbin Institute of Technology, China
[e] Faculty of Design Architecture and Building, University of Technology Sydney, Australia

ARTICLE INFO

Keywords:
Automated
Hazard identification
Zone classification
Density maps

ABSTRACT

Current onsite safety management always relies on time-consuming predefinitions of hazardous zones based on the managers' personal capabilities. However, in a typical labor-intensive industry such as construction, the workers themselves can provide a wealth of information for hazard identification. Historical accident-free working locations on site provide a valuable means of recognizing safe workplaces. This paper presents an approach to the automated classification of construction site zones derived from the location tracks of workers collected from a real-time location system (RTLS). Through data mining, filtering and analysis, the location tracks are transformed into grid density maps and continuous density maps. These illustrate the characteristics of spatial-temporal activities onsite as well as providing a visual representation of the distribution of safe and hazardous individual workplaces. A personnel hazard map is generated automatically based on historical accident-free location tracks from a field project using the proposed approach. Compared with the actual workplaces in terms of accuracy, precision, sensitivity and specificity, the evaluation result reveals that the hazardous areas on a construction site can be automatically classified to improve the workplace management of individual workers. The contributions of this research include an automated zone classification algorithm and an evaluation framework consisting of four indicators for hazard awareness onsite.

1. Introduction

Identifying the changing hazards or controlling risks during construction activities is an important, but often quite difficult task. This is especially the case onsite, even when most of the activities are conducted repetitively [1], where it is almost impossible to avoid all the safety hazards in a workspace due to the complex nature of construction projects [2,3]. Construction sites are also highly dynamic, with exposed workspaces and their occupation constantly changing, exacerbating the already serious hazard identification problem for both the construction site and crew. Since it is uneconomical or ineffective to employ more safety inspectors, an efficient and automated approach is needed.

There is an increasing use of personal mobile devices integrated with geographic location tracking, context-awareness and wireless communication in the construction industry, and working habits are changing accordingly. This is providing the potential for accessing a wealth of information for evaluation, communication and collaboration onsite. Basic data concerning the continuously changing locations of communication devices onsite enables the geographic position and spatial-temporal behavior of workers, materials and equipment to be monitored by simple manipulation, providing managers and workers with opportunities for creative initiatives for the collection, tracking and visualization of onsite construction activities [4–9]. This has given rise to the introduction of location-based services (LBS) that offer value-added services for individuals in the form of new utilities embedded in their personal devices [10]. Properly leveraged, this rich spatial-temporal information has the potential to improve hazard identification and the control of risks onsite.

However, the increasing amount of research applying LBS to safety issues is mostly based on predefined unalterable manual rules. Proximity hazard indicators between workers, equipment and hazardous areas are widely employed in pro-active real-time construction systems due to the growing body of evidence indicating that potential

* Corresponding author at: ZN1002, BRE Dept., Z Core, Phase 8, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong.
E-mail addresses: heng.li@polyu.edu.hk (H. Li), xincong.yang@outlook.com (X. Yang), rm.skitmore@qut.edu.au (M. Skitmore), wflai@sina.com (F. Wang), Perry.Forsythe@uts.edu.au (P. Forsythe).

http://dx.doi.org/10.1016/j.autcon.2017.04.007
Received 17 August 2015; Received in revised form 21 March 2017; Accepted 6 April 2017

risks and accidents can be reduced by avoiding working in, or close to, a dangerous location at a specific time [11–16]. The workspace requirements of labor and equipment operations in 3D BIM models are being generated with increasing precision to improve the efficacy of proximity alert systems and approaches [17,18]. Although these studies greatly assist in safety management onsite, pre-construction safety plans are still insufficiently adequate to cope with the dynamic and multiple objectives that occur during daily onsite activities. The whole hazard identification process needs to be involved prior to the application of these proximity approaches [2], otherwise, unidentified hazards will continue to threaten the health and safety of the workforce. Moreover, the situation is exacerbated by most planners being conservative and unable to provide timely updates of hazardous locations according to changing site conditions.

Since the construction process lasts so long, it is reasonable to consider the hazard zones to be static at short time intervals. Accordingly, this study aims to develop an automated approach to identifying, mapping and updating all of the area-restricted hazards or safe zones onsite in a timely manner. This involves deriving crucial historical locations deemed to be safe working zones, such as accident-free walk paths, by *crowd sourcing* (workers engaged in similar activities or in the same group) to assist in individual safety decision-making [19]. This exploratory approach attempts to classify the entire site into hazardous and safe zones through a novel *peer based* approach based on their frequency of occupation by workers, potentially providing an available means to reuse the historical data for further prediction in the short run. On the assumption that areas that have been occupied by accident-free workers are more likely to be safe areas than otherwise, the issue then becomes one of identifying such areas. The approach utilizes data mining and information technology to extract and integrate density maps from these areas to provide individual guides to safe zones in the form of personal hazard maps [20].

Consequently, to achieve the objectives of this paper that harness accident-free work trajectories and safety preferences by like-minded peers, the rest of the paper is structured as follows. Section 2 investigates the background of the research, containing traditional hazard identification approaches and potential issues in practice. Then the core proposed framework consisting of four modules is introduced in Section 3. Curial zones of workplaces are visualized via density maps to display their distribution and mark the characteristics of workers onsite. In Section 4, a field case study is described to demonstrate the capability to create an automated zone classification map for an individual worker and evaluate its accuracy. Finally, conclusions and future research possibilities are provided in the last section.

2. Background

Although the associated root causes of fatal/serious accidents are well known, including lack of attention, insufficient safety training, tiredness, poor quality materials and equipment [21], there nonetheless still exist unidentified hazards or risks that cannot be anticipated prior to their occurrence. From an external environment perspective, the hazards are a result of a variety of circumstances, including unexpected site conditions. The constantly changing dynamic of aggregated variables onsite also undermines hazard identification [2]. From an internal worker perspective, different workers share different safe and hazardous zones due to human factors [22,23]. For instance, a ditch on a site may be a safe working zone for an experienced excavator operator who understands the work method involved and uses appropriate personal protective equipment (PPE), but may be a hazardous zone for other workers. Both perspectives make it impossible to identify all the hazards involved completely in advance.

Commonly, most accidents onsite are regarded as the result of contact collisions mainly caused by low awareness and blind spots [12,24]. Thus, apart from site inspections, proximity safe alert systems based on real-time location systems (RTLS) have been extremely popular and unsafe-proximity identification is widely used to provide proactive safety management [1]. Spatial interference between personnel, related equipment and materials, such as the proximity of labor to operating heavy equipment or moving vehicles, can be detected or predicted [5,13], For example, based on onsite dynamics, the analysis of activities and related hazards, Guo has identified space conflicts by considering space constraints and path interferences to assist decision-making [25]; Sacks et al. has designed an algorithm to estimate the likelihood of spatial and temporal exposure to related hazards [1]; Lee et al. have developed a radio frequency identification (RFID)-based RTLS suitable for diverse sites to contribute to aggressive safety management [16]; and Marks and Teizer propose proximity detection between workers and equipment [12]. To enhance the efficiency of proximity safe systems, Kim et al. have developed a human-assisted obstacle avoidance system during equipment operation [26]; Wang and Razavi have constructed a low false alarm rate model by adding position, heading direction and speed attributes [15]; and Cheng et al. further propose to utilize the fusion of RTLS and physiological status as well as thoracic posture to activity analysis [27,28]. On the other hand, Vahdatikhaki and Hammad, Tantisevi and Akinci generate a dynamic equipment workspace [17,29,30]; Akinci et al. design a project-specific model to build workspace requirements at the activity-level [31]; with Zhang et al. then integrating BIM into the 3D visualization of the workspace requirements to promote the accurate calculation of proximity [18]. These studies all require the manual pre-identification of unsafe-proximity prior to applying field-testing, which is time-consuming and prone to invalidate the approaches due to unpredicted conditions.

Since it is impossible and uneconomical for managers to identify all the unsafe-proximity hazards before construction, a novel approach to extracting spatial-temporal information from historical location tracks is to use the wealth of information available of the locations of workers, materials and equipment [32]. This can be conveniently obtained by utilizing such advanced technology as UWB, RFID and GPS [4,33–35]. A few researchers have attempted to promote proximity safe alert systems through learning from the spatial-temporal proximity relationships of near misses. Wu et al., for example, have designed an autonomous system by considering the characteristics of near misses based on typical historical accident cases [24]; and Teizer and Cheng have collected and studied near-miss data to provide a proximity hazard indicator to identify obstacles for route searching and generate heat maps for safety planning [36,37], which enhances safety knowledge sharing among stakeholders [13]. These studies attempt to transfer safety knowledge from not only inspectors and managers but also the workforce itself. However, such approaches commonly view and serve workers as a community, with often-insensitive unsafe-proximity recommendations being provided for safety initiatives.

Therefore, an approach that automatically updates according to movement feedback from specific groups of workers onsite has a genuine potential to provide the responsiveness required sufficiently and efficiently. However, a major problem with historical location resources is that enormous amounts of information have to be examined in order to find the relevant pieces needed. If the working area of a worker is safe, then it should also be safe for similar workers undertaking similar activities at a similar time. Thus, integrating the density maps of historical accident-free locations should provide a much better alternative than increasing the number of safety managers onsite.

3. Framework of automated zone classification

The framework of the proposed approach is represented in Fig. 1. The origin dataset is cloud-stored and contains historical locations, real-time locations, layout and predefined special zones. It is not compulsory to input the data in the dot boxes since sometimes workers are new to the site or adequate detailed information of the geographic attributes of sites cannot be obtained before construction. The essential assumptions

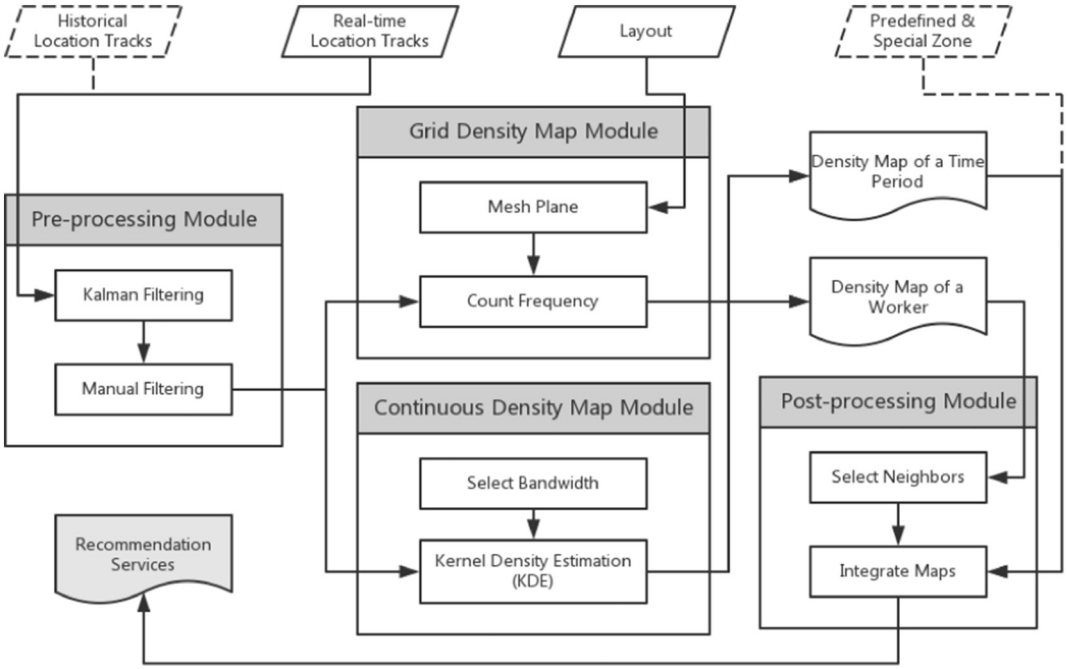

Fig. 1. Framework for automated zone classification.

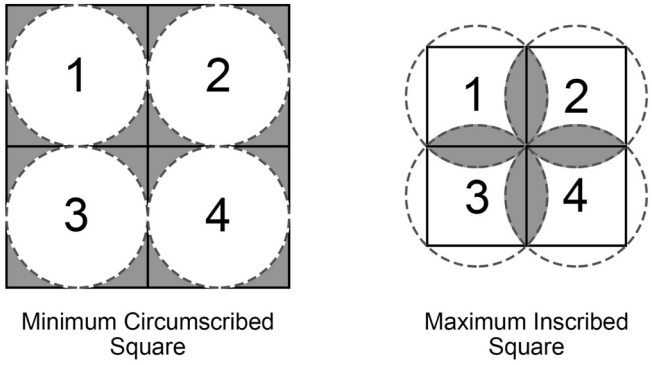

Fig. 2. MCS and MIS grid.

of the framework are first described and each module is illustrated in the remainder of this section.

3.1. Essential hypotheses

Two essential hypotheses are proposed in order to provide the foundation for the processing modules. That is:

Hypothesis 1. Workers generally tend to avoid hazards instinctively. While workers are working or moving from one workplace to another, they tend to walk around hazards at an appropriate distance [2]. Hence, an accident-free location can involve one of two possible circumstances: (1) the location is completely within a safe, hazard-free area; or (2) the location is partially within a hazardous area where the associated hazards are under control. Taken in short time intervals, a dynamic construction site can be considered static and thus such areas can be treated as temporal safe areas with only a tiny probability of accidents.

Hypothesis 2. Workers always undertaking the same or similar tasks onsite face similar hazards and risks. Though many factors, such as training levels, health conditions, etc., have an obvious impact on workers' decisions and actions, to simplify the process here only the task is taken into account in zone classification as the same concepts hold for other factors involved in the task. Therefore, workers in the same task group are treated as a crowd source sharing the same safe zones.

3.2. Data collection

The core dataset required to be input contains three parts, the real-time locations, working characteristics of labor crew and the layout of the project. The first part can be collected from a real-time location system based on either of RFID, UWB, WLAN, GPS etc. [33–35]. These systems require the manual deployment of equipment and tags ahead of normal operation. For example, the RTLS used in this study is developed based on a wireless personal area network (WLAN) [42] in which the pre-installation of sensing infrastructure and tags are essential before actual construction. Generally, numerous workers carry out activities last for a long time per day and therefore the overall amount of locations is so large that processing with ordinary computation capability onsite is problematic. To remedy this, the entire location arrangement is simplified and only part of the locations is taken into account. The available components in this study are the locations where workers are involved in direct and hauling activities, which are identified by observers in the following pre-processing module. Note that a satisfactory degree of accuracy in a Cartesian coordination representation is also necessary for further analysis. The second part - the working characteristics of workers - containing working experience, task allocation and relevant team members, can be determined by observing and investigating the subjects directly or by referring to the managers daily reports. The last part is layout, which can be derived from shop drawings or aerial photography from drones, providing the current situation of the construction site.

3.3. Pre-processing module

The objective of the pre-processing module is to filter out and smooth incorrect location boundaries. Although off-the-shelf hardware customized to suit the needs of a precise location can be used, the signals received may still contain unwanted noise hidden inside the dataset [37], leading to a location with cusps and occasional outliers. Since this noise is mainly the result of measurement rather than processing errors, a high proportion follows a normally distribution. The Kalman filter, as a robust algorithm, is therefore used to eliminate

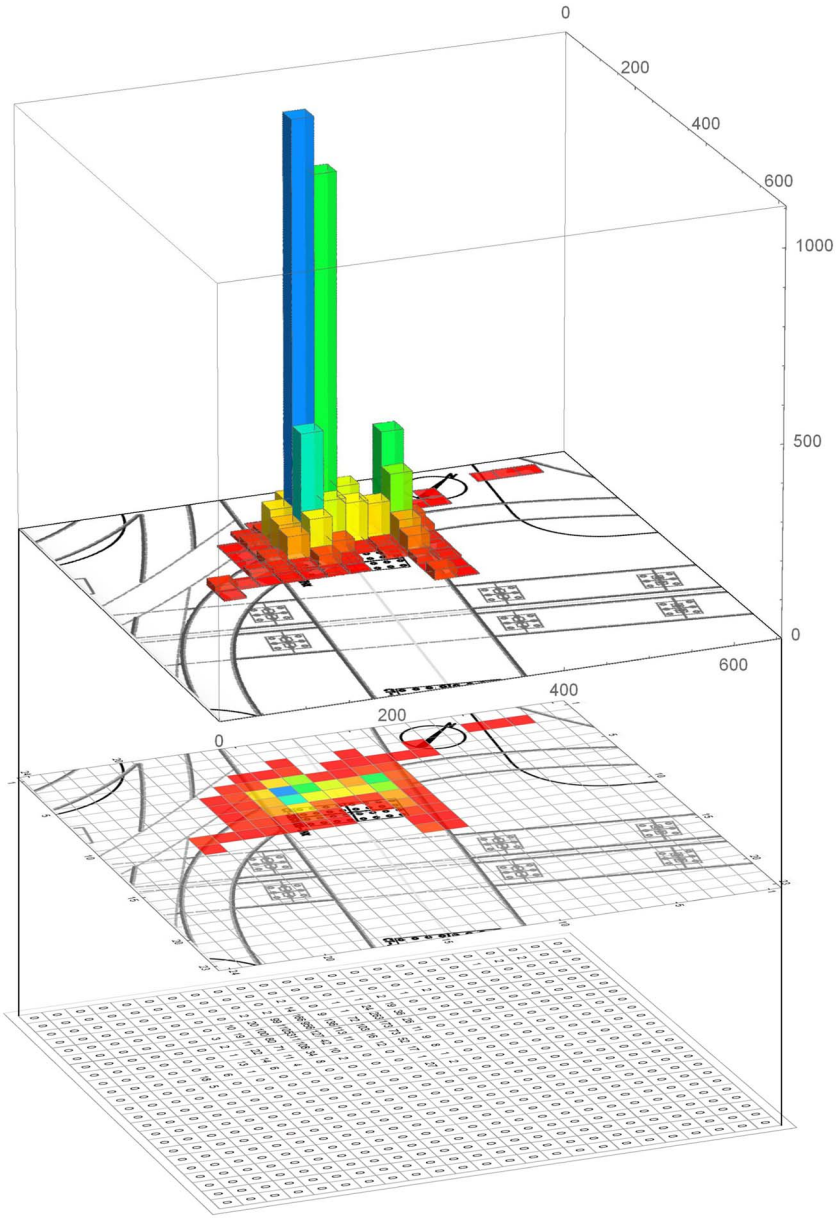

Fig. 3. Conversion of density matrix and grid density map.

the noise and smooth the location boundaries [38]. Compared with the extended and federated Kalman filter, the linear Kalman filter is likely to be more effective and easier to apply in real-time situations [39,40].

The following manual filter aims to overcome two kinds of site-related problems. The first concerns the initialization of the system, in that abnormal signal readings might be received while initializing the configuration settings. The other concerns the discontinuity of tasks occurring on the site. The intervals between activities include the workers' breaks or rest times and the inevitable waiting time, such as in the hardening time of concrete. Workers might relax outside the detectable zones or log out leading to duration intervals. Because the focus of the proposed framework aims to analyze the main workplaces, these irrelevant records have to be eliminated manually. For inaccurate initialization locations, an inference variable is added into the RTLS in the form of a fixed point onsite. The location signals of workers are ignored until the inference signals are constantly received. For irrelevant locations from rests, etc., an upper bound is determined by observation or feedback once the rest time is over the threshold, so the locations during the rest are ignored.

3.4. Grid density map module

In order to extract workplace information and establish the fundamental characteristics of the workers, the whole construction site is divided into grids by constructing a two-dimensional lattice layout for analysis. This module comprises two steps: (1) meshing the layout plane and (2) counting the frequency of occupancy in each grid and drawing the density maps.

3.4.1. Mesh plane

Assume that the Pro-Active Construction Management System (PCMS) [41] is used as the main RTLS. This is designed to send and receive signals every 0.5 s with the broadcast frequency of around 2 Hz. Thus, the influence area of a single worker is a circle with a radius of 1 m, assumed to be the potential moving distance in a half second. For simplicity and fast segmentation, the grid shape of the influence area is transformed into a square. Thus, the dimensions of the maximum inscribed square (MIS) and the minimum circumscribed square (MCS) are 1.414 m and 2 m respectively. As Fig. 2 illustrates, if the detection points are located in the gray areas, the points for the MIS are regarded

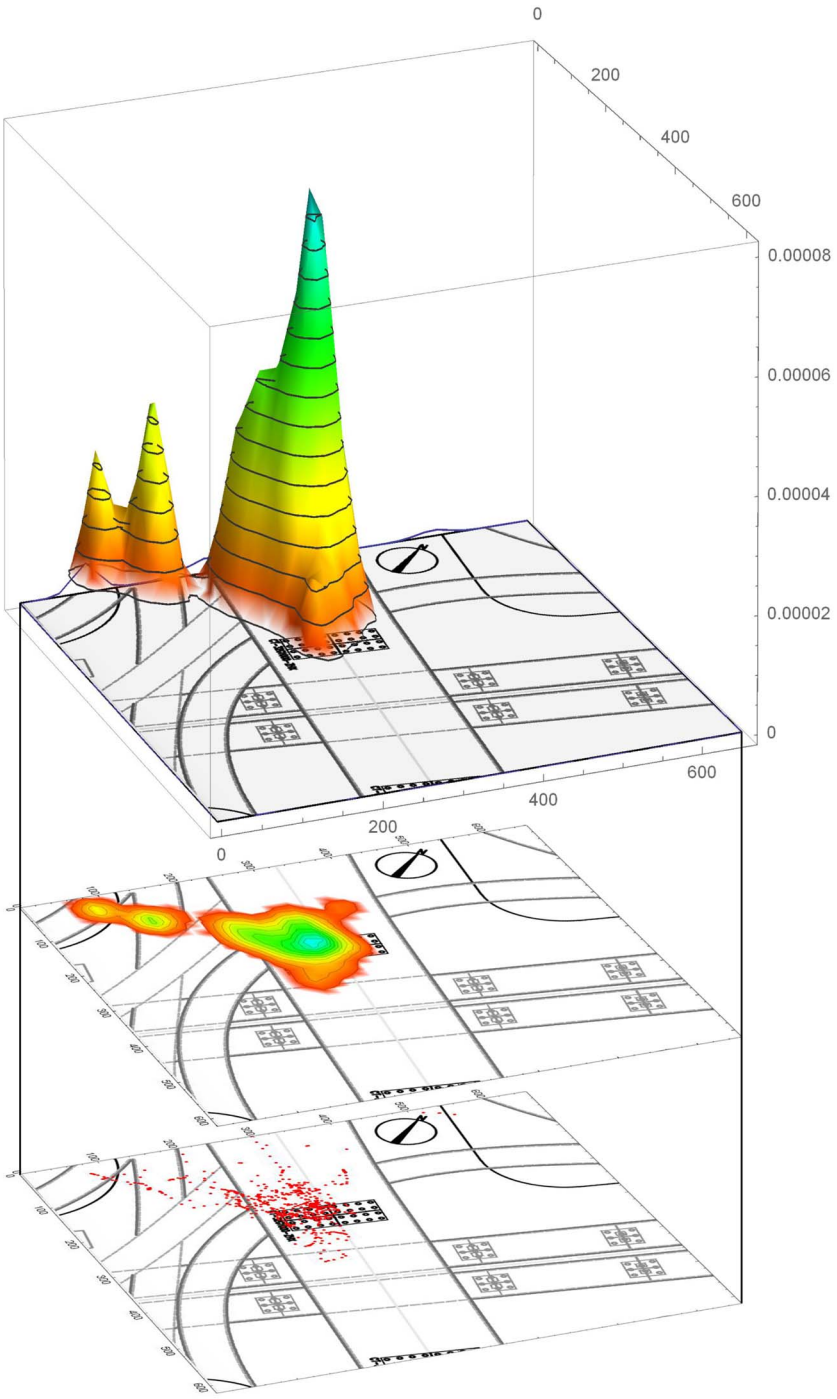

Fig. 4. Transformation to a continuous density map.

as in the grid but not actual influence areas, with an error rate of 0.214; while the points for MCS are regarded as in another grid instead of the actual one, with an error rate of 0.429. Thus, compared with the 9 m × 9 m grid for earthmoving operations [42], a 3 m × 3 m grid for trucks [43] and 0.5 m × 0.5 m grid for workplace requirements analysis [18], the whole site is divided into a 2 m × 2 m chequerboard.

3.4.2. Count frequency

To describe the characteristics of the spatial distribution of workers and activities, density maps as the common thematic maps can offer a comprehensive picture of the actual construction site [44]. A form of discrete and abrupt density map of existing spaces between observations – the grid density map (also called a heat map) is employed in this module to quantitatively deal with the tracked locations. Just as the histogram provides a graphic visualization of one-dimensional numerical data, the grid density map represents the distribution of two-dimensional movement data. When the construction site is divided into a series of equal grids, known as bins, the algorithm counts how many movement points fall into each grid and enters the frequency into the density matrix to plot the grid density map as shown in Fig. 3.

3.5. Continuous density map module

Considering the complex nature of construction sites, a qualitative description might be sometimes used to simplify the reasoning process dealing with inaccurate location tracks as well as to show the crucial

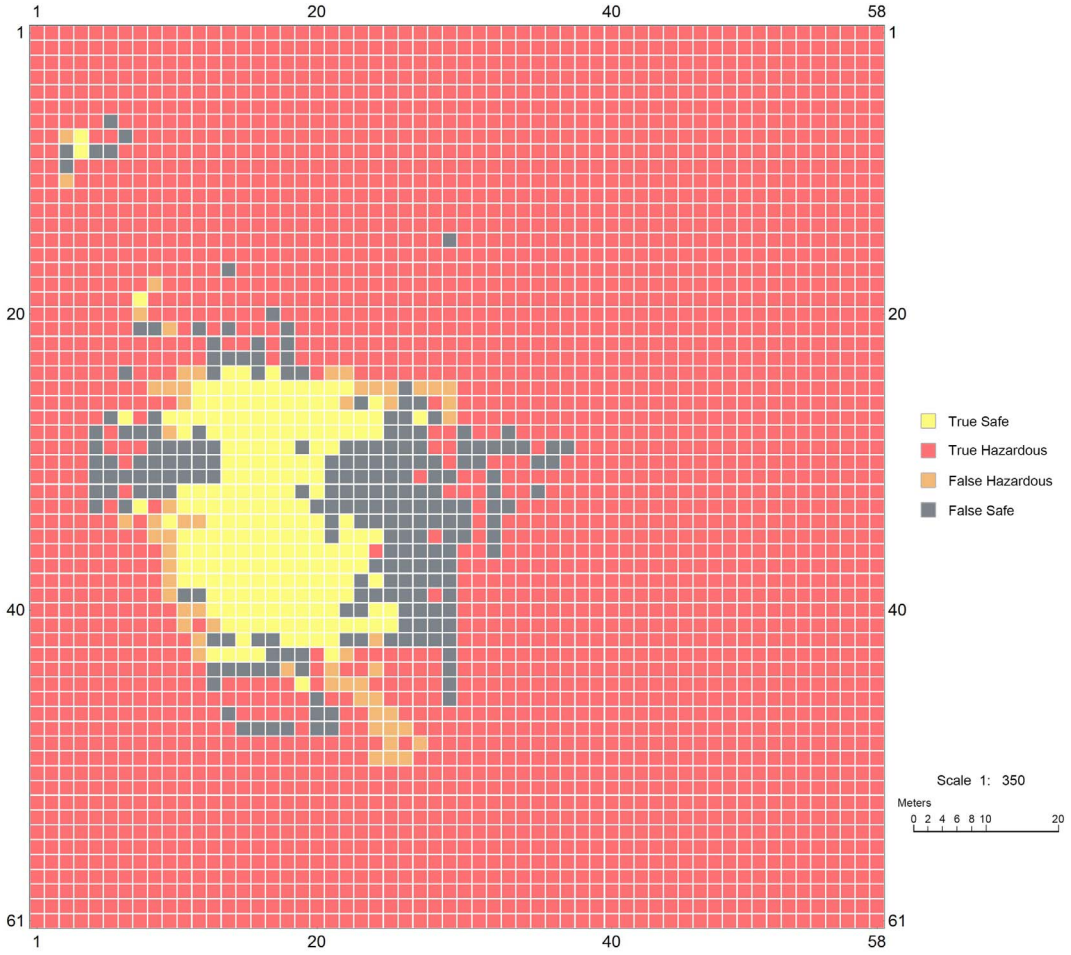

Fig. 5. Grid density map of a sample worker.

factors involved. Thus, this module contains additional two-steps for plotting: a kernel density estimate (KDE) and bandwidth selection.

3.5.1. Kernel density estimation (KDE)

The grid density map can be treated as a two-dimensional histogram with grid length and width, which is simple and user-friendly. However, the actual distribution might be skewed if parts of the location tracks are lost. In particular, different orientations and grid dimensions can also result in discrepancies in the visualization of the same distribution. KDE is an efficient algorithm to alleviate these issues, making the plots smoother and more robust.

As a popular non-parametric density estimator, KDE has various kernel functions, including Epanechnikov, Biweight and Cosine. Most are radially symmetric and unimodal, such as the Gaussian kernel function widely employed, with

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$$

and the corresponding kernel estimator for one-dimensional data is

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)$$

where n is the number of samples and the bandwidth, and h is the smoothing parameter. However, location tracks are multivariable cases, and features with disproportionate weights in need of a covariance matrix can result in a complicated procedure. In the present case, however, the dimensions of location tracks - x and y - are independent, sharing a same kernel function. Therefore, the product kernel is used here as a suitable alternative for multivariate KDE, as

$$\hat{f}(x, y) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h_x h_y} K_x\left(\frac{x - X_i}{h_x}\right) K_y\left(\frac{y - Y_i}{h_y}\right)$$

3.5.2. Bandwidth selection

Bandwidth is a crucial factor affecting the slopes in KDE. Too large a bandwidth over-smooths the density estimation and disguises the structure of the data, while a bandwidth that is too small under-smooths the density estimation and forces it to be too spiky to interpret. Therefore, an optimal bandwidth is selected to minimize the combination of bias (system errors) and variance (random errors) as well as to cover all the features simultaneously. The mean integrated squared error (MISE) is the standard optimality criterion, where

$$\mathrm{MISE}(h) = E \int \left(\hat{f}_h(x) - f(x)\right)^2 dx$$

If the true distribution is assumed Gaussian, the optimal bandwidth as initial input to fulfill the optimality criterion approximates to

$$h_{\mathrm{Gaussian}}{}^{*} \approx 1.06 \sigma n^{-\frac{1}{5}}$$

where σ is the sample standard deviation and n is the sample size. However, in practical field tests, the hidden distribution of trajectory is unknown. The initial bandwidth requires updating using a data-based method such as Gaussian approximation or Silverman's rule.

Consequently, continuous density maps can be drawn as shown in Fig. 4 based on the estimated function to identify the features of workplaces.

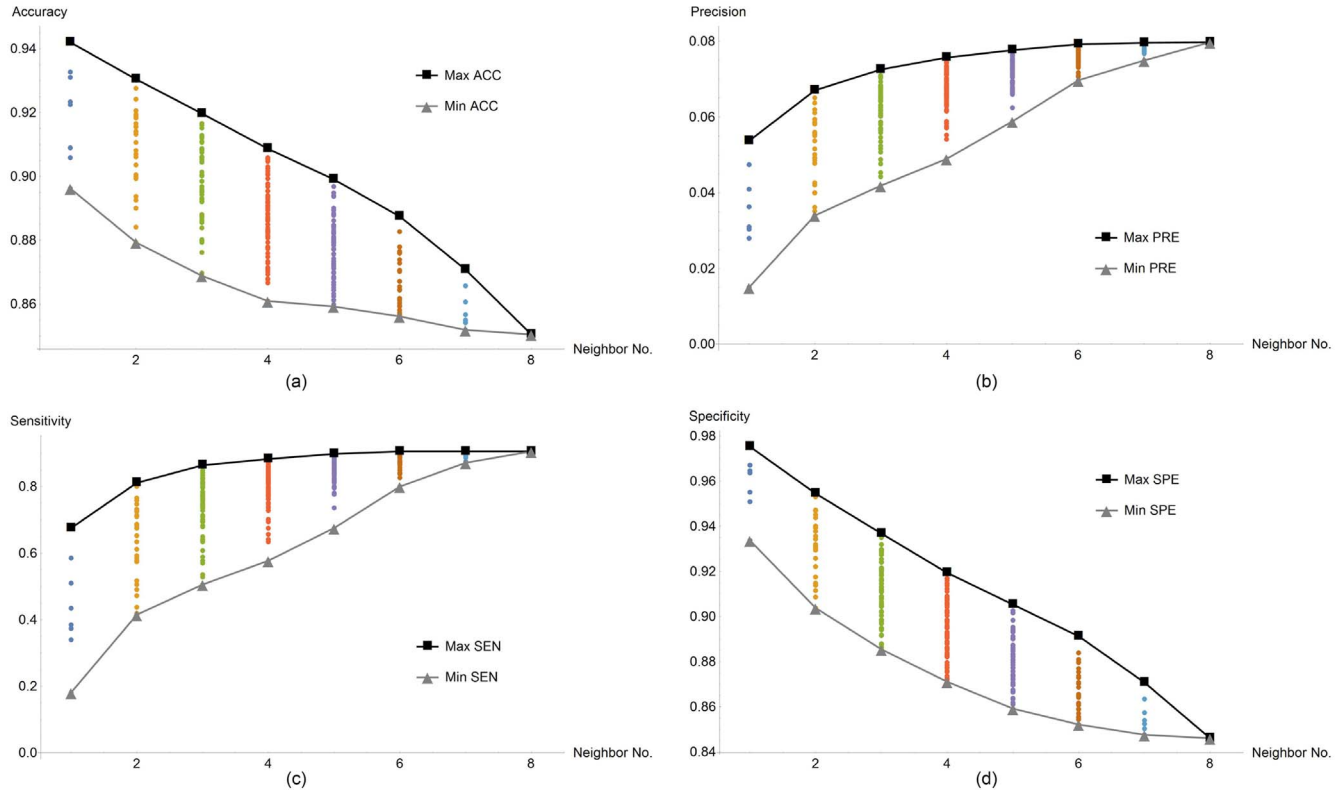

Fig. 6. Analysis of the four indicators with increasing number of neighbors.

3.6. Post-processing module

This module deals with the grid and continuous density maps generated by the former modules to fulfill different zone classification requirements. For personal zone services, density maps of relevant workers are selected and integrated to guide the identification of the worker safety zones. For population zone services, on the other hand, density maps of a certain time period are transformed into a sequence showing the dynamics on site. The current approach focuses on formal safety recommendations.

3.6.1. Neighbor selection

Neighbors are defined here as like-minded peers, such as a co-operator in a team, among an onsite crowd source and who are engaged in the same, similar or related activities. This subset of the workforce commonly experiences major threats posed by the same hazards. For instance, being aware of a hazardous floor hole, a bar bender attempts to carry out construction tasks at a distance from the hole. As a result, no location tracks within the hazardous area can be derived from RTLS. Other bar benders also only accept the workplace without the hole as being safe, while the rest of the workforce, such as excavators whose workplace might contain the hole, would not consider the hole to be a serious hazard. Therefore, the current accident-free locations of workers are potentially safe working zones. In this exploratory research, all the tracked workers are aggregated into data sets by occupation.

3.6.2. Integrate maps

To accomplish density map integration, each worker is treated as having the same impact on other similar neighboring workers. Thus, the density maps of neighbors can be superimposed linearly on a layout or actual site map following the superposition principle.

For grid density maps, the integrated procedure can be formulated as a weighted mixture.

$$g(x,y) = \sum_{i \in N} w_i f_i(x,y)$$

where x and y represent the row and column of the grid perspective, N is the dataset of neighbors, w_i denotes the weights corresponding to neighbor i decided by variables such as experience and $f_i(x,y)$ is the discrete function of the count frequency. Here, all weights are taken as equal to each other without considering personal attributes such as age or gender.

For continuous density maps that qualitatively illustrate the distribution of the main workplaces, the ultimate zone classification indicating safety probability is referred to as a mixture density. Since the number of peer workers is finite and countable, the density sum is the same formula as for the grid density maps. However, an extra step needed in practice is that the data-based bandwidth has to be reselected in order to combine or separate the distribution kurtosis, with a flatter or more pointed peak and slighter or heavier tails.

Area recommendation services would ultimately be carried out based on the integrated maps throughout the framework. Grid density maps provide precise safety and hazard guides in which grids with high frequency are treated as potential safety zones and vice versa. The continuous density maps indicate the important workplaces upon which to focus and re-filter out abnormal grids, showing the dynamics of the construction site and changing workplaces. Managers can then check the zone classification and mark other predefined special zones on the final maps.

4. Case study

To examine the approach, assume that a group of workers involved in a bridge project are tracked by a location tracking technology, a series of historical location tracks are obtained with certain accuracy that satisfies the requirements of density computing. Here, the hazards on construction sites were mainly comprised of electrical and fire dangers; and the experienced workers tended to conduct construction

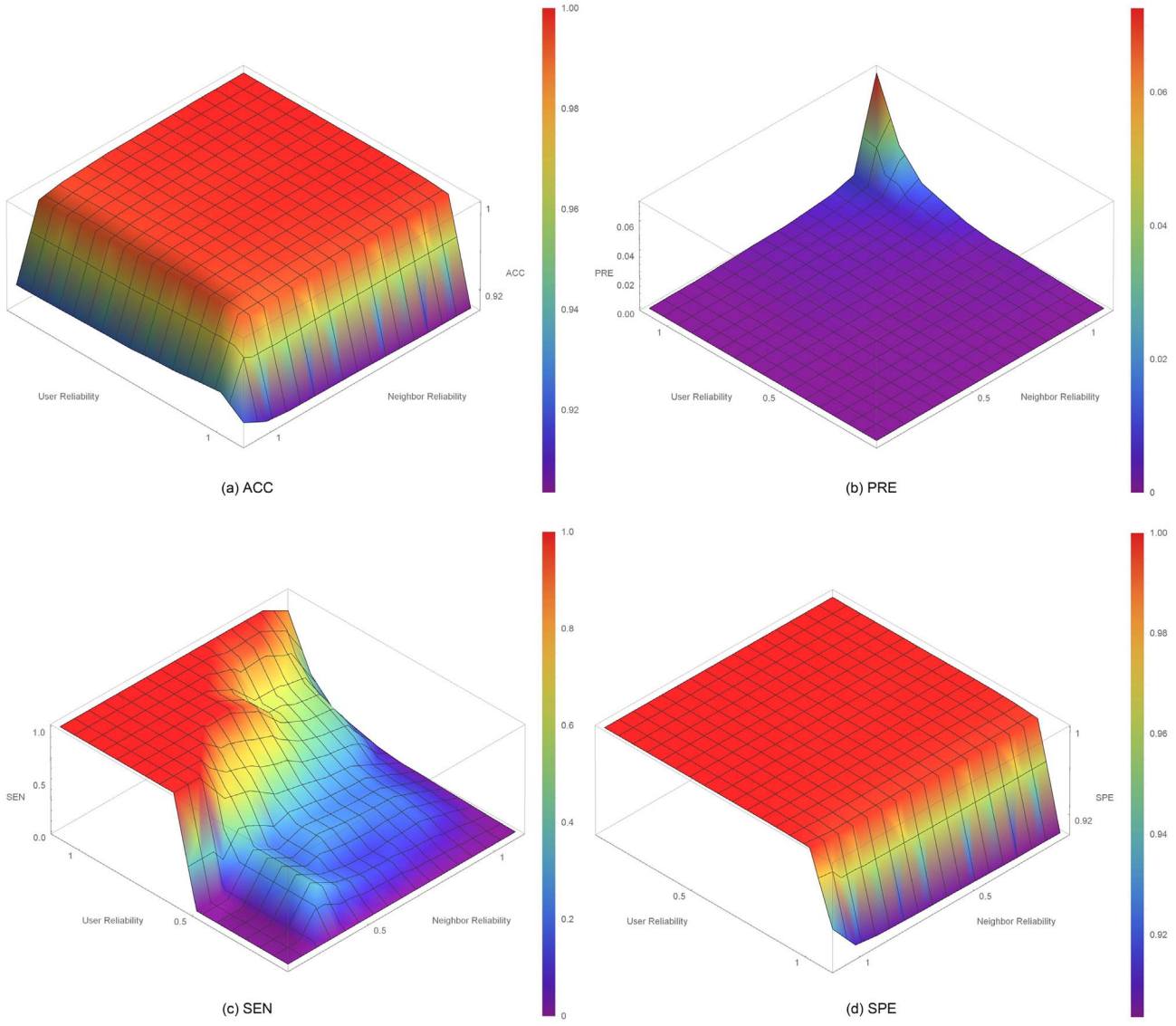

Fig. 7. Analysis of the four indicators of three neighbors by their reliability ratio.

activities far away from these hazards. Thus, accident-free location data were firstly pre-processed with a Kalman filter and a manual filter to smooth and eliminate the obvious erroneous points due to accidental errors, and the dimension of grids was determined to be 2 m × 2 m according to literature. The density maps were then drawn based on the occupancy of each grid by individual workers. A worker was selected randomly to be a target worker, while the remaining workers were selected by random sampling to comprise the neighborhood set for the target worker. Compared the safety and hazard zones of target worker with the generated density maps from neighborhood set, the proposed approach could be examined and validated. To provide significant help to layout planners, operators and managers, the system needs to identify core hazardous areas accurately with an economic warning distance. This means that the accuracy, precision, sensitivity and specificity of the safe and hazardous workplace classifications are crucial. Here, *accuracy* measures the bias of the core hazard and safety locations between the actual workplace and density maps; *precision* indicates the variability of the zone classification, seriously impacting on the boundaries; *sensitivity* assesses the performance of classification by measuring the proportion of safety zones that are correctly identified; and *specificity*, which can be regarded as an economic indicator, refers to the probability of encountering actual hazards in all the hazardous zones. To measure the efficacy of the approach, the actual accident-free locations of any worker are selected as the reference for comparison with the zone classification based on the crowd source subset.

4.1. Validation of grid density maps

The integrated density maps are transformed into a binary matrix according to the frequency of the grids. As Fig. 5 shows, the potential outcomes are defined as:

True safe (*TS*): safe grids correctly identified as safe.
False safe (*FS*): hazardous grids incorrectly identified as safe.
True hazardous (*TH*): hazardous grids correctly identified as hazardous.
False hazardous (*FH*): safe grids incorrectly identified as hazardous.

The four indicators are calculated by

$$\text{Accuracy (ACC)} = \frac{\text{number of } TS + TH}{\text{number of } TS + FS + TH + FH}$$

$$\text{Precision (PRE)} = \frac{\text{number of } TS}{\text{number of } TS + TH}$$

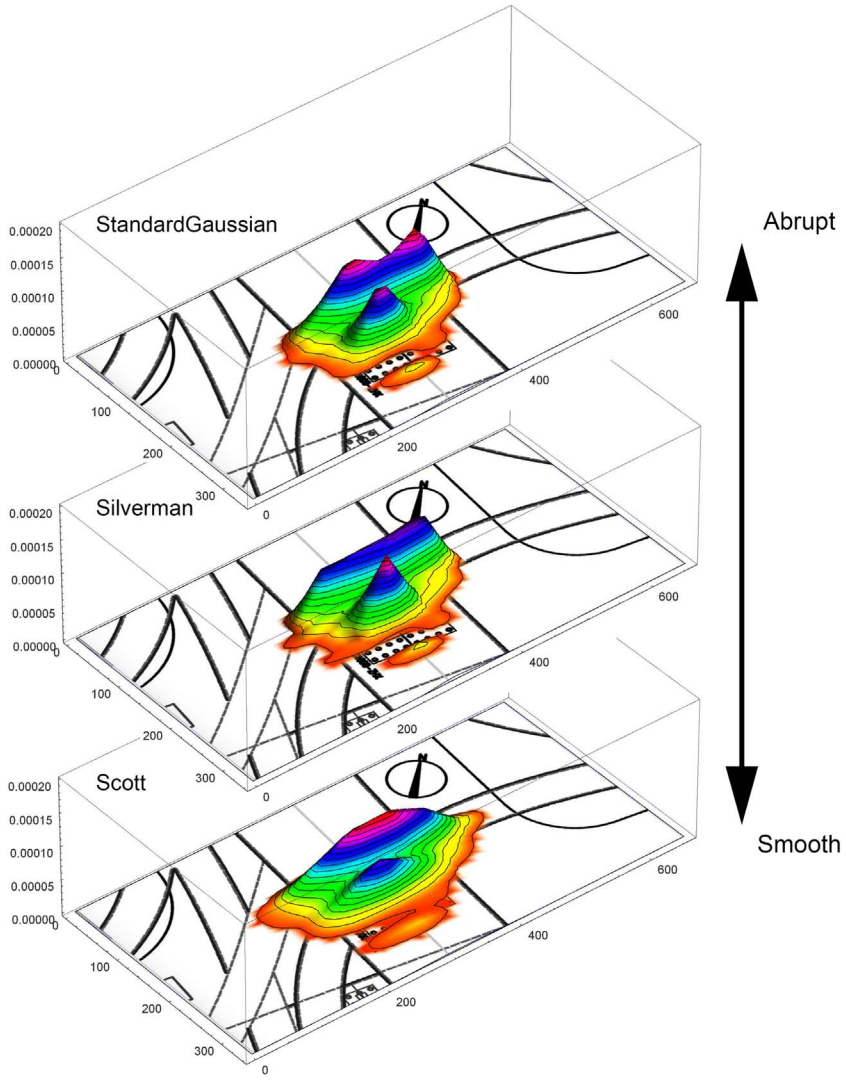

Fig. 8. Bandwidth selection.

$$\text{Sensitivity (SEN)} = \frac{\text{number of } TS}{\text{number of } TS + FH}$$

$$\text{Specificity (SPE)} = \frac{\text{number of } TH}{\text{number of } TH + FS}$$

If the fellow workers are all experienced and cautious, the grids for the accident-free locations can be assumed to indicate a completely safe workplace. The results of the four indicators for a bar bender are presented in Fig. 6. Apart from the target worker, the eight fellow workers comprising the neighbor set provide 255 combinations of subsets. Although both accuracy and specificity decline gradually with an increasing number of neighbors, their ranges are still over 80%. One possible explanation is that the decreasing homogeneity of the workers involved in the reference set resulted in the rapid expansion of safe zones so that the proportion of potentially hazardous areas for the target worker, but safe areas for fellow workers, would rise. Although the precision was relatively low due to the small proportion of hazards onsite, there was a dramatic increase in sensitivity. This might be attributed to the robustness of the crowd resource, in that the involvement of more workers would eliminate the individual zone characteristics.

However, historical location tracks, such as in the preliminary period of new projects, are not easily accessible in practice. Therefore, the safety probability of the collected location tracks is added as an argument to enhance reliability. Commonly, the working locations of experienced workers provide more reliable information than new workers and therefore they are allocated a higher proportion in map integration. Considering the computing requirements involved and the total size of samples, three neighbors were selected to be temporal references as they provided acceptable evaluated indicators.

As Fig. 7(a) illustrates, when the reliability ratio of the neighbors and target worker falls, there is a sharp rise in accuracy to almost 1, especially for the edge areas. This perhaps indicates that eliminating a few frequency grids significantly increases accuracy. Since system errors are inevitable and lead to some abnormal location tracks, using reliability ratio filtering at low-frequency locations could improve the effectiveness of the system. By contrast, precision falls suddenly with the reliability ratio, which is interpreted as due to the reduction of available grids from neighbors and user, leading to less mutual workplaces. Thus, the amount of TS grids gradually decreases while TH rapidly increases. The graphs displayed in Fig. 7(c) and (d) show that sensitivity and specificity tend to develop asymmetrically when considering reliability ratios. Lower reliability ratios result in lower sensitivity and higher specificity, especially for the user, since the individual workplaces are always so small that reliability has a profound effect in comparison with neighbors.

This evaluation illustrates the influence of the number of neighbors and reliability on system performance and demonstrates the validity of the grid density maps. Although the indicators are not stable and are seriously affected by the combinations of workers selected and their

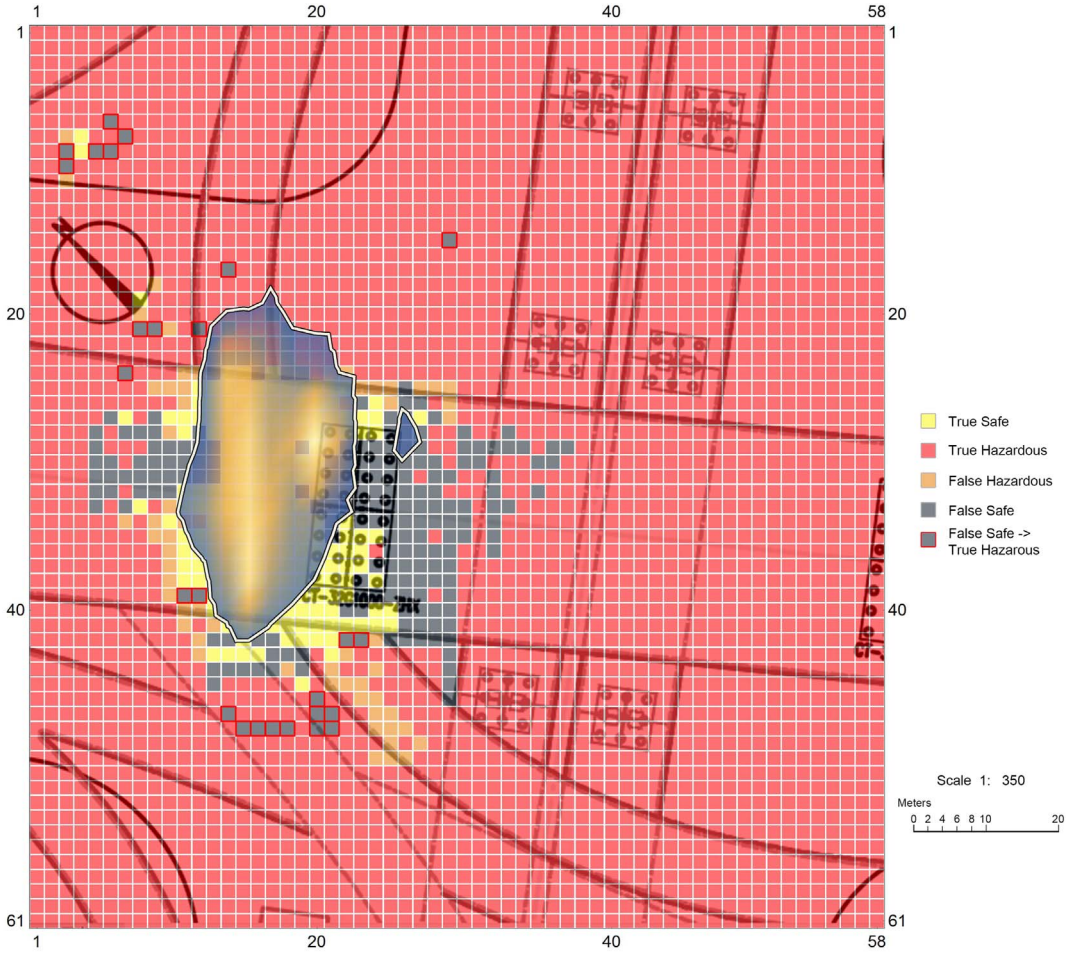

Fig. 9. Integrated density map of a sample worker.

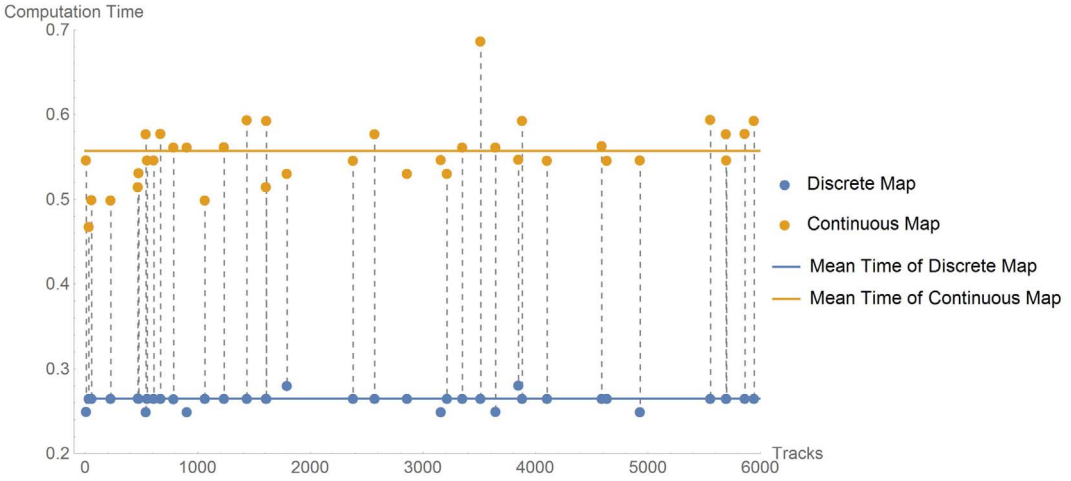

Fig. 10. Computation time for generating discrete and continuous density maps.

working experience and dynamic changes in frequency on site, the approach nevertheless has considerable potential for identifying most safe and hazardous zones onsite.

4.2. Validation of the continuous density maps

Unlike density maps, continuous density maps assist in zone classification by taking into account the population distribution onsite. As a crucial factor in the performance and appearance of KED methods, bandwidth is selected first based on the same three neighbors in the above case. The 3-D surfaces in Fig. 8 show the conceptualized distribution based on three bandwidth selection methods: Standard Gaussian, Silverman's rule and Scott's rule. Although all identify the main workplaces of the three neighbors, the Standard Gaussian does this in an abrupt fashion, while Scott's rule combines and flattens the distribution, expanding the predicted safety zones and leading to an inaccurate horizontal projected area. Therefore, Silverman's rule is selected as being the most suitable.

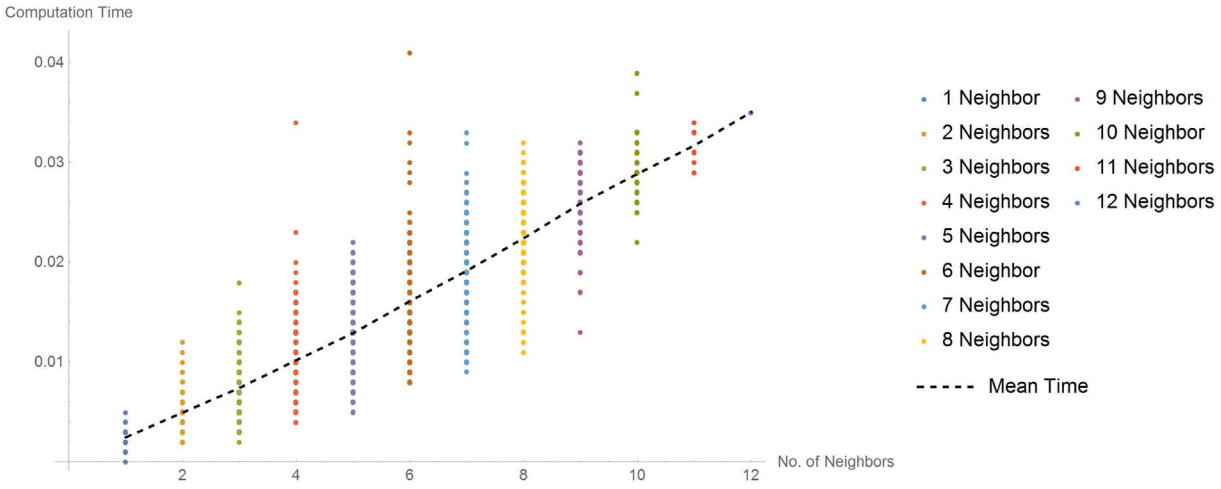

Fig. 11. Computation time of computing the density matrix from neighbors.

Since construction workplaces are spatial-temporal continuous, the continuous density maps show the possible distribution as actual field practices. Although the continuous density maps have fuzzy boundaries, they assist in reinforcing true grids and eliminating false grids. The fragmentary grids outside the distribution without any connection to core workplaces are likely to be incorrect owing to system error and therefore need to be filtered out.

Fig. 9 shows the integrated grid density map and continuous density map overlaid on the planned construction site layout. The white closed line represents the contour plot of the assumed distribution of neighbors. Based on the boundaries of the continuous density maps, the sparse grids outside the normal distribution of the workplaces of neighbors are, as with the grids surrounded with thick boxes, eliminated owing to abnormal tracks. Therefore, the previous FS grids that are classified as safe but are actually hazardous, are transformed into a TH grid. Compared with the single grid density map classification, the integrated method increases accuracy and specificity from 91.97% and 93.17% to 92.71% and 93.24% respectively, decreasing precision from 6.27% to 6.22% and maintaining sensitivity at 76.98%.

Hence an analysis of algorithm complexity was performed by computation time for generation of discrete and continuous density maps with respect to location amount, which was operating on an HP Z400 workstation with Intel Xeon W3550 (3.06 GHz). As Fig. 10 reveals, the computation time, including CPU computing and plotting, holds stable values with the increase in the number of location tracks. Obviously, the computation time for continuous density maps was higher than that for discrete 2 m × 2 m grid maps due to the complexity of KDE. Both the absolute times take an average of less than 1 s, suggesting that the update of zone classification can be made in a timely manner. Another Fig. 11 shows the core computation time for generating discrete density maps with the respect to neighbor set size. With the number of neighbors increasing, the mean computation time increases linearly. The calculation time is reasonable since the map integration is linear and simple. However, because of the discrepancies in location tracks, the computation time ranges are changing unsteadily.

Although not all hazards were identified and cold-starts could not be accommodated (due to the lack of historical records of a new site or location tracks), once the cloud dataset of the workers was established, the hazardous locations onsite could be recognized automatically. Throughout the case study, the final integrated density maps were evaluated to be effective and valuable for managers and for individual workers' better personal safety on site.

5. Conclusions, limitations and future research

Manual hazard identification for construction safety management is time-consuming and difficult to accommodate site dynamics. This paper presents a novel approach to zone classification, leveraging the historical working locations of peer crowd-source workers. After preprocessing, the location tracks from the RTLS are delivered to two core modules to extract the grid density maps and continuous density maps respectively. The whole site is divided into grids to calculate the frequency of occupancy of locations, with KDE concurrently estimating the possible distribution of the locations. Finally, density maps of a time period or worker are integrated to provide a zone classification guide.

The approach contributes to enhancing hazard identification onsite without the need for predefinition or utilizing knowledge and experience in recognizing safe/hazardous workplaces. The automated zone classification extracted from accident-free locations is made by computational data mining and processing, making use of the historical tracking capacity of the PCMS. The maps generated also offer a novel form of visualization of spatial-temporal characteristics or workplaces and their possible distribution. Indicators of accuracy, precision, sensitivity and specificity computed by comparing binary matrices in a novel way are used to evaluate the performance of the approach.

A case study is used for demonstrating, testing and validation, and shows how the changing evaluation indicators with reliable ratios and bandwidth selection can be used to enhance performance. Although not all reference workers are experienced and whose main working locations cannot always be treated as safe workplaces, the approach to zone classification is shown to be feasible and provides a wealth of information concerning the spatial-temporal distribution of workers onsite. Using grid density maps and continuous density maps with geographical data onsite, the case study indicates the potential of the approach to act as a very accurate automated workplace safety guide.

However, it is noticed that this approach cannot support the identification of all hazardous areas exactly onsite. That could be attributed to the assumptions made for the proposed model as well as equipment limitations. One of the major arguments was that not all the hazards onsite are static and interrelated with locations, such as truck crashes, which was not tested by the case project but is common in practice. Furthermore, personal training levels, health conditions, protective equipment, etc., might significantly affect workers' decisions and behaviors, which could not be taken into account in this study because of the size of teams. In addition, cold-starts and hazardous areas smaller than a grid could not be accommodated or identified due to the lack of historical records of a new site or tiny regions. Such problems could only be solved if the cloud dataset of the workers was established and real-time data was input to the model using feedback loops, leading to a more realistic and automated prediction. With respect to equipment limitations of accuracy, the RTLS used on realistic construction sites still requires improvement to perform better and

more reliably. Future research needs to utilize advanced technologies and measurements, such as iBeacon and laser scanning, to improve the location accuracy in a comprehensive outdoor environment, as well as consider more variables, to enhance the effectiveness of zone classification. In addition, future studies are needed to propose a framework to select crowd source automatically to enhance the current algorithm.

Acknowledgement

We gratefully appreciate the funding support from Research Grants Council of Hong Kong with the grant entitled “Proactively Monitoring Construction Progress by Integrating 3D Laser-scanning and BIM” (PolyU 152093/14E).

References

[1] R. Sacks, O. Rozenfeld, Y. Rosenfeld, Spatial and temporal exposure to safety hazards in construction, J. Constr. Eng. Manag. 135 (8) (2009) 726–736, http://dx.doi.org/10.1061/(ASCE)0733-9364(2009)135:8(726).

[2] H. Kim, H.-S. Lee, M. Park, B. Chung, S. Hwang, Automated hazardous area identification using laborers' actual and optimal routes, Autom. Constr. 65 (2016) 21–32, http://dx.doi.org/10.1016/j.autcon.2016.01.006.

[3] G. Carter, S.D. Smith, Safety hazard identification on construction projects, J. Constr. Eng. Manag. 132 (2) (2006) 197–205, http://dx.doi.org/10.1061/(ASCE)0733-9364(2006)132:2(197).

[4] A.R. Andoh, X. Su, H. Cai, A framework of RFID and GPS for tracking construction site dynamics, Proceedings of the ASCE Construction Research Congress (CRC), 2012, http://dx.doi.org/10.1061/9780784412329.083.

[5] N. Pradhananga, Construction Site Safety Analysis for Human-equipment Interaction Using Spatio-temporal Data (Doctoral Dissertation), (2014).

[6] S.N. Razavi, C.T. Haas, Multisensor data fusion for on-site materials tracking in construction, Autom. Constr. 19 (8) (2010) 1037–1046, http://dx.doi.org/10.1016/j.autcon.2010.07.017.

[7] J. Song, C. Haas, C. Caldas, Tracking the location of materials on construction job sites, J. Constr. Eng. Manag. 132 (9) (2006) 911–918, http://dx.doi.org/10.1061/(ASCE)0733-9364(2006)132:9(911).

[8] J. Teizer, T. Cheng, Y. Fang, Location tracking and data visualization technology to advance construction ironworkers' education and training in safety and productivity, Autom. Constr. 35 (0) (2013) 53–68, http://dx.doi.org/10.1016/j.autcon.2013.03.004.

[9] C. Zhang, A. Hammad, H. Bahnassi, Collaborative multi-agent systems for construction equipment based on real-time field data capturing, Journal of Information Technology in Construction 14 (Special Issues on Next Generation Construction IT: Technology Foresight, Future Studies, Roadmapping, and Scenario Planning), 2009, pp. 204–228.

[10] C. Ratti, S. Williams, D. Frenchman, R. Pulselli, Mobile landscapes: using location data from cell phones for urban analysis, Environ. Plann. B. Plann. Des. 33 (5) (2006) 727, http://dx.doi.org/10.1068/b32047.

[11] J. Teizer, B.S. Allread, C.E. Fullerton, J. Hinze, Autonomous pro-active real-time construction worker and equipment operator proximity safety alert system, Autom. Constr. 19 (5) (2010) 630–640, http://dx.doi.org/10.1016/j.autcon.2010.02.009.

[12] E.D. Marks, J. Teizer, Method for testing proximity detection and alert technology for safe construction equipment operation, Constr. Manag. Econ. 31 (6) (2013) 636–646, http://dx.doi.org/10.1080/01446193.2013.783705.

[13] J. Teizer, T. Cheng, Proximity hazard indicator for workers-on-foot near miss interactions with construction equipment and geo-referenced hazard areas, Autom. Constr. 60 (2015) 58–73, http://dx.doi.org/10.1016/j.autcon.2015.09.003.

[14] F. Vahdatikhaki, A. Hammad, Visibility and proximity based risk map of earthwork site using real-time simulation, Proceedings of the 32nd International Symposium on Automation and Robotics in Construction (ISARC), 10 2015.

[15] J. Wang, S.N. Razavi, Low false alarm rate model for unsafe-proximity detection in construction, J. Comput. Civ. Eng. (2015) 04015005, http://dx.doi.org/10.1061/(ASCE)CP.1943-5487.0000470.

[16] H.-S. Lee, K.-P. Lee, M. Park, Y. Baek, S. Lee, RFID-based real-time locating system for construction safety management, J. Comput. Civ. Eng. 26 (3) (2011) 366–377, http://dx.doi.org/10.1061/(ASCE)CP.1943-5487.0000144.

[17] F. Vahdatikhaki, A. Hammad, Dynamic equipment workspace generation for improving earthwork safety using real-time location system, Adv. Eng. Inform. (2015), http://dx.doi.org/10.1016/j.aei.2015.03.002.

[18] S. Zhang, J. Teizer, N. Pradhananga, C.M. Eastman, Workforce location tracking to model, visualize and analyze workspace requirements in building information models for construction safety planning, Autom. Constr. 60 (2015) 74–86, http://dx.doi.org/10.1016/j.autcon.2015.09.009.

[19] J. Herlocker, J.A. Konstan, J. Riedl, An empirical analysis of design choices in neighborhood-based collaborative filtering algorithms, Inf. Retr. 5 (4) (2002) 287–310, http://dx.doi.org/10.1023/A:1020443909834.

[20] T. Suzuki, D. Miyoshi, J. Meguro, Y. Amano, T. Hashizume, K. Sato, J. Takiguchi, Real-time hazard map generation using small unmanned aerial vehicle, SICE Annual Conference, 2008, 2008, pp. 443–446, , http://dx.doi.org/10.1109/SICE.2008.4654695.

[21] E.W. Cheng, H. Li, D. Fang, F. Xie, Construction safety management: an exploratory study from China, Constr. Innov. 4 (4) (2004) 229–241, http://dx.doi.org/10.1108/14714170410815114.

[22] P. Forsythe, Proactive construction safety systems and the human factor, Proceedings of the ICE-Management, Procurement and Law, 167 (5) 2014, pp. 242–252, , http://dx.doi.org/10.1680/mpal.13.00055.

[23] C. Fraser, The influence of personal characteristics on effectiveness of construction site managers, Constr. Manag. Econ. 18 (1) (2000) 29–36, http://dx.doi.org/10.1080/014461900370924.

[24] W. Wu, H. Yang, D.A.S. Chew, S.-h. Yang, A.G.F. Gibb, Q. Li, Towards an autonomous real-time tracking system of near-miss accidents on construction sites, Autom. Constr. 19 (2) (2010) 134–141, http://dx.doi.org/10.1016/j.autcon.2009.11.017.

[25] S.J. Guo, Identification and resolution of work space conflicts in building construction, J. Constr. Eng. Manag. 128 (4) (2002) 287–295, http://dx.doi.org/10.1061/(ASCE)0733-9364(2002)128:4(287).

[26] C. Kim, C.T. Haas, K.A. Liapi, C.H. Caldas, Human-assisted obstacle avoidance system using 3D workspace modeling for construction equipment operation, J. Comput. Civ. Eng. 20 (3) (2006) 177–186, http://dx.doi.org/10.1061/(ASCE)0887-3801(2006)20:3(177).

[27] T. Cheng, G. Migliaccio, J. Teizer, U. Gatti, Data fusion of real-time location sensing and physiological status monitoring for ergonomics analysis of construction workers, J. Comput. Civ. Eng. 27 (3) (2013) 320–335, http://dx.doi.org/10.1061/(ASCE)CP.1943-5487.0000222.

[28] T. Cheng, J. Teizer, G.C. Migliaccio, U.C. Gatti, Automated task-level activity analysis through fusion of real time location sensors and worker's thoracic posture data, Autom. Constr. 29 (0) (2013) 24–39, http://dx.doi.org/10.1016/j.autcon.2012.08.003.

[29] K. Tantisevi, B. Akinci, Automated generation of workspace requirements of mobile crane operations to support conflict detection, Autom. Constr. 16 (3) (2007) 262–276, http://dx.doi.org/10.1016/j.autcon.2006.05.007.

[30] F. Vahdatikhaki, A. Hammad, Risk-based look-ahead workspace generation for earthwork equipment using near real-time simulation, Autom. Constr. 58 (2015) 207–220, http://dx.doi.org/10.1016/j.autcon.2015.07.019.

[31] B. Akinci, M. Fischer, J. Kunz, Automated generation of work spaces required by construction activities, J. Constr. Eng. Manag. 128 (4) (2002) 306–315, http://dx.doi.org/10.1061/(ASCE)0733-9364(2002)128:4(306).

[32] H. Kim, H.-S. Lee, M. Park, B. Chung, S. Hwang, Information retrieval framework for hazard identification in construction, J. Comput. Civ. Eng. 29 (3) (2013) 04014052, http://dx.doi.org/10.1061/(ASCE)CP.1943-5487.0000340.

[33] A. Oloufa, M. Ikeda, H. Oda, GPS-based wireless collision detection of construction equipment, NIST Special Publication SP, 2003, pp. 461–466.

[34] A. Rodriguez, C. Zhang, A. Hammad, Feasibility of location tracking of construction resources using UWB for better productivity and safety, The Int'l Conference on Computing in Civil and Building Engineering, 2010, p. 2010 (Nottingham, United Kingdom).

[35] J. Song, C.T. Haas, C.H. Caldas, A proximity-based method for locating RFID tagged objects, Adv. Eng. Inform. 21 (4) (2007) 367–376, http://dx.doi.org/10.1016/j.aei.2006.09.002.

[36] O. Golovina, J. Teizer, N. Pradhananga, Heat map generation for predictive safety planning: preventing struck-by and near miss interactions between workers-on-foot and construction equipment, Autom. Constr. (2016), http://dx.doi.org/10.1016/j.autcon.2016.03.008.

[37] T. Cheng, U. Mantripragada, J. Teizer, P.A. Vela, Automated trajectory and path planning analysis based on ultra wideband data, J. Comput. Civ. Eng. 26 (2) (2012) 151–160.

[38] S.N. Razavi, C.T. Haas, Reliability-based hybrid data fusion method for adaptive location estimation in construction, J. Comput. Civ. Eng. 26 (1) (2012) 1–10, http://dx.doi.org/10.1061/(ASCE)CP.1943-5487.0000101.

[39] Z.M. Durovic, B.D. Kovacevic, Robust estimation with unknown noise statistics, IEEE Trans. Autom. Control 44 (6) (1999) 1292–1296.

[40] P. Katharina, Linear Kalman Filter for Position and Orientation Tracking, Technische Universität München, Technical Report, 2005.

[41] R.E. Kalman, A new approach to linear filtering and prediction problems, J. Fluids Eng. 82 (1) (1960) 35–45, http://dx.doi.org/10.1115/1.3662552.

[42] H. Li, M. Lu, G. Chan, M. Skitmore, Proactive training system for safe and efficient precast installation, Autom. Constr. 49 (Part A) (2015) 163–174, http://dx.doi.org/10.1016/j.autcon.2014.10.010.

[43] N. Pradhananga, J. Teizer, Cell-based construction site simulation model for earthmoving operations using real-time equipment location data, Visualization in Engineering, 3 (1) 2015, pp. 1–16, , http://dx.doi.org/10.1186/s40327-015-0025-3.

[44] C. Zhang, A. Hammad, T.M. Zayed, G. Wainer, H. Pang, Cell-based representation and analysis of spatial resources in construction simulation, Autom. Constr. 16 (4) (2007) 436–448, http://dx.doi.org/10.1016/j.autcon.2006.07.009.