Association for Information Systems

AIS Electronic Library (AISeL)

Quantity over Quality? – A Framework for Combining Mobile Crowd Sensing and High Quality Sensing

Barbara Stöckel
FZI Forschungszentrum Informatik, Karlsruhe, Germany

Simon Kloker
Karlsruhe Institute of Technology, Karlsruhe, Germany

Christof Weinhardt
Karlsruhe Institute of Technology, Karlsruhe, Germany

David Dann
Karlsruhe Institute of Technology, Karlsruhe, Germany

Follow this and additional works at: https://aisel.aisnet.org/wi2021

Quantity over Quality? – A Framework for Combining Mobile Crowd Sensing and High Quality Sensing

Barbara Stöckel[1], Simon Kloker[2], Christof Weinhardt[2], David Dann[2]

[1] FZI Forschungszentrum Informatik, Karlsruhe, Germany
stoeckel@fzi.de
[2] Karlsruhe Institute of Technology, Karlsruhe, Germany
{simon.kloker, christof.weinhardt, david.dann}@kit.edu

Abstract. Mobile Crowd Sensing is a widespread sensing paradigm, successful through the ever-growing availability of mobile devices and their increasing sensor quality. Mobile Crowd Sensing offers low-cost data collection, scalability, and mobility, but faces downsides like unknown or low sensing quality and uncertainty about user behavior and movement. We examine the combination of traditional High Quality Sensing methods and Mobile Crowd Sensing in a Hybrid Sensing system in order to build a value-creating overall system, aiming to use both sensing methods to ensure high quality of data, yet also benefiting from the advantages Mobile Crowd Sensing has to offer such as mobility, scalability, and low deployment cost. We conduct a structured literature review on the current state and derive a classification matrix for Hybrid Sensing applications.

Keywords: Mobile Crowd Sensing, High Quality Sensing, Data Combination, Hybrid Sensing, Design Science.

1 Introduction

By now, *Mobile Crowd Sensing (MCS)* is a widespread large-scaled sensing paradigm. However, some of the most prominent applications, building on information gathered via MCS, like the traffic prediction service embedded in Google Maps, which counts more than one billion active users per month [1], belong to the subcategory of "opportunistic crowdsensing" [2]. This means data is shared without active user intervention and, consequently, is often not perceived as MCS by the user. MCS approaches belonging to the other subcategory of "participatory crowdsensing" [3], where users actively contribute data or information via mobile devices, have received more scientific attention so far. Although these applications do not generate revenue on the same scale as Google, many of them still have a notable number of active users. The user-generated database of Pl@ntnet [4], for instance, contains almost 1.8 million images covering nearly 28 thousand species [5] and does not only help plant enthusiasts to identify their uploaded plant photos automatically, but also makes an active contribution to preserve protected areas [6].

The spread of MCS is predominantly driven by the increasing presence of mobile devices (e.g., smartphones) as well as improved integrated sensor technology and their

increasing performance (e.g., accuracy, battery life). In addition, the expansion of advanced mobile internet technologies (e.g., 4G, 5G) enables a variety of innovative new MCS applications, especially those depending on real-time processing, as large amounts of data can be transferred to the cloud with virtually no time delay.

In contrast, *High Quality Sensing (HQS)* describes traditional methods to collect data (or information), guaranteeing an almost error-free recording of the variable to be surveyed (e.g., via sensor, expert observation). These professional measurements, sometimes gathered within industrial contexts, are of high quality and verified. However, as they are disadvantageous in terms of mobility, scalability, deployment, and maintenance, they face the problem of being expensive and lacking in spatial and temporal coverage [7].

As MCS and HQS have, to a certain extent, complementary strengths and weaknesses, combining both methods in a *Hybrid Sensing (HS)* system offers the possibility to exploit the strengths of each method in order to compensate for the weaknesses of the other. While MCS comes with advantages in terms of mobility, causing improved spatial-temporal coverage and low-cost scalability [7], [8], as well as the ability to provide additional information for better context awareness through human input [9], HQS can provide reliable, high-quality data [10], [11] available for quality improvement of the overall system, for instance via sensor calibration or training of prediction models applicable on MCS data. Additionally, a cost-reduction may be achieved through optimal resource allocation in terms of energy consumption [12], maintenance or sensor deployment [13], when combining both sensing methods.

Against this backdrop, there is a variety of scientific literature that focuses on the combination of MCS and HQS. However, existing approaches differ substantially with regard to different aspects (e.g., strategic focus, data type) and there is no structured overview or general approach, resulting in a lack of knowledge transfer to other application areas.

Consequently, we argue it is time to take a step back and assess the current state of affairs by structuring existing approaches, methods and results. Our research question is therefore:

RQ1: How can existing approaches on the value-creating combination of MCS and HQS be classified?

We approach this RQ by conducting a structured literature review, in which we condense the insights of 23 papers, which feature (prototypically) implemented and theoretical HS approaches. Since the results show that a structured approach towards building an HS system is still missing, we decided to address this striking research gap as well. Therefore, our second research question is:

RQ2: What are the essential components when designing a HS system that combines MCS and HQS in value-adding way?

In a parallel and dependent process to deriving a classification for existing HS approaches (answering RQ1), we iteratively develop a HS framework (answering RQ2). These two tasks represent the main work of the first of three cycles in our overarching Design Science (DS) project [14], within which we aim to build a generic value-creating process for combining MCS and HQS. Whilst the focus of this paper lies on the results of the first cycle, in future work we will build a prototypical implementation

of our HS framework (cycle two), for which we have chosen the context of road condition monitoring (RCM). We will leverage the prototype's evaluation results to further refine and generalize our approach later on (cycle three).

RCM is a suitable context for our research endeavour in so far, as the HS applications and theoretical approaches that we identified in literature primarily focus on stationary high-quality sensors (e.g., [10], [11]), leaving the application of mobile high-quality sensors, such as necessary in RCM, as a clear research gap. Conventional high-quality approaches on RCM include scanning the road profile with a high-quality Lidar sensor and determining the International Roughness Index (IRI), a global measurement for longitudinal evenness. This sensing method is expensive, due to the needed equipment and personnel expenses, and, therefore, not feasible for ensuring a high spatial-temporal coverage. The road condition on German federal motorways, for instance, is only recorded at fixed intervals of four years [15]. Moreover, roads in the federal states, districts, and municipalities are excluded and thus subject to individual local maintenance plans, leading to inconsistent road assessment. In recent years, various smartphone-based solutions [16] were implemented using data collected via smartphone sensors (e.g., accelerator, gyroscope) to predict the road quality (e.g., RoadSense [17], Roadroid [18]). Although they help to increase the spatial-temporal coverage, they cannot guarantee a reliable high quality. They do not attempt to include the structured and continuous combination of both MCS and HQS data in a HS system, which we therefore address as a novelty in our RCM use case (in cycle two), in order to solve the quality deficit.

The contribution of this paper is three-fold. First, with our structured literature review, we identify and structure existing approaches, methods, and findings. To the best of our knowledge, our paper represents the first structured literature review on this timely and important topic. Second, we introduce a novel HS classification matrix, which logically groups approaches with similar major tasks and challenges, enabling the classification of HS applications and consequently facilitating the transferability of knowledge. Third, we present the first version of a HS framework, which we will further refine in DS cycle two and three, which aims to provide a generalized approach to building a HS system for fellow researchers and practitioners.

The remainder of this paper is structured as follows. In the next section, we describe our research project and the DS methodology used. Section 3 reports on our results from the structured literature review on combining MCS and HQS. Based on the structured literature review, we then derive two interim artifacts (HS classification matrix and HS framework) in Section 4, providing more insight into our RCM use case. Section 5 discusses our results and paths for future research.

2   Design Science Research Methodology

In our research project, we employ a Design Science Research Methodology following the guidelines of Kuechler and Vaishnavi [14], which excels by its strong emphasis on an iterative procedure in rapid iterating cycles, making the development of the artifact flexible in its ability to react to re-evaluated requirements. Although several quick

iterations are conducted, we define three main cycles for providing a contextual structure. In these three main cycles, we aim to build a HS framework, for value-creating combination of MCS and HQS, with a prototype implementation in a RCM use case using mobile high-quality sensors, which represents a novelty in the domain. With our DS project, we seek to solve the lack of a structured cross-disciplinary approach when it comes to the combination of MCS and HQS, due to which the potential offered by HS is not fully exploited and the transfer of derived knowledge to other applications is restrained. Our HS framework aims to help fellow researchers and practitioners to design HS applications and improve the knowledge transfer to their respective fields of application.

In the *first cycle*, we aim to clarify the problem space and classify different types of HS applications. Based on the knowledge deducted from literature, we draft an overall HS framework, according to whose design we will implement a prototype RCM application in the second and third cycle. Since the literature gives only incomplete information about the required components of the HS framework, the implementation in the following two cycles serves for evaluation and further refinement based on the results. The first cycle is already fully completed, the results of which are the main focus of this paper. In the *second cycle*, the operational infrastructure for the RCM application is set up, a crowd app is developed and extensively tested. An exploratory data analysis, data cleaning, and pre-processing steps for the MCS and HQS data are performed as a basis for the following data combination and model training. The working steps described up to this point have already been carried out following the

General DS Cycle	Design Cycle One HS classification and framework	Design Cycle Two RCM instantiation	Design Cycle Three Refinement and generalisation
Awareness of Problem	Structured literature review on combining MCS and HQS to identify problem and research gaps	Refinement of problem identification RCM problem identification	(Refinement of problem identification)
Suggestion	Derivation of meta-requirements (relevant components) for HS framework	Revision of meta-requirements for HS framework	Revision of meta-requirements for HS framework
Development	Derivation of HS classification matrix Development of HS framework	Instantiation of HS system in RCM application	Revision and generalisation of HS framework Revision of RCM implementation
Evaluation	HS classification matrix and HS framework draft are iteratively deducted from literature and continuously revised	Evaluation of HS framework based on RCM instantiation Expert testing to evaluate functionality of RCM application	Field test for quantitative evaluation Expert workshop to evaluate utility and derive Design Principles
Conclusion	Result: Hybrid Sensing Classification Matrix	Result: Initial applicable RCM HS system	Result: Design Theory (HS Framework)

Figure 1. Design Science Research Methodology by Kuechler and Vaishnavi [14] applied to our project with three main cycles (executed tasks are highlighted in color, development in cycle two is currently in process)

first results of cycle one. Building on this, we will implement and evaluate a combination approach based on our HS framework in future work. This will comprise the geospatial data fusion (Data Combination), model training for predicting the road quality based on sensor data (Data Processing) and system quality assessment (Quality Evaluation). The HS framework drafted in the first cycle and continuously refined in the second will be evaluated and discussed in the light our specific use case. In the *third cycle*, we want to use the results from cycle one and two to revise the overall HS framework and its instantiation in the RCM use case. We will further use the enhanced RCM application to derive specific conclusions, which we aim to generalize, regarding the overall HS framework including the topics of spatial-temporal coverage and quality measures, as well as identifying relevant quality enrichment tasks (e.g., task allocation, incentives) for improving the HS system.

3 Literature Review

In the following section, we present the results of our structured literature review, which we conduct as a central task of the first design cycle. We thereby raise awareness of the problem and create the basis for the derivation of the preliminary artifacts described in more detail in Section 4.

3.1 Approach to Literature Review

We conduct a structured literature review following the methodological suggestions by Webster and Watson [19] and vom Brocke et al. [20]. In an explorative search, we identify an initial pool of literature for extracting relevant keywords, based on which we build a search term[1] for our structured literature search. We query a set of interdisciplinary databases (i.e., ACM Digital Library, AIS eLibrary, Emerald Insight, IEEE Xplore Digital Library, ProQuest, ScienceDirect/Scopus, Web of Science) for matching our search term in title, abstract, or keywords [20]. After removing duplicates, this results in 134 publications for further review. Since there is no commonly used term for what we refer to as HQS, we use describing and closely related terms for it in our search query resulting in 118 irrelevant findings, that are not concerned with the combination of MCS and HQS, which we identify by analyzing title and abstract. With the remaining 16 papers as our initial pool of literature, we conduct a successive backward and forward search, which result in further seven relevant publications, yielding a total of 23 articles. The identified articles both contain theoretical approaches, focusing on architecture or simulated models, and practical approaches with concrete or prototype implementations. We consider the paper [7] to be relevant as it adequately outlines opportunities and challenges of HS, but does not contain a HS

---

[1] ("mobile crowdsensing" OR "mobile crowd sensing" OR "participatory sensing") AND ("industrial sens*" OR "traditional sens*" OR "stationary sens*" OR "static sens*" OR "special* sens*" OR "sensor node*" OR "expert contribut*" OR "industrial data" OR "hybrid" OR "industrial IOT" OR "industrial Internet of Things") *Note:* * represents one or more wildcard characters

approach. We present the results of our literature review grouped into classes, the derivation of which we explain in detail in Section 4.1, in the context of the development of the HS classification matrix.

In the following two subsections, we first clarify the terms MCS and HQS and then present an overview over existing HS approaches grouped in classes, highlighting the focus of the work.

3.2 MCS and HQS: Clarification of Terms

Guo et al. [21] defines MCS as "a new sensing paradigm that empowers ordinary citizens to contribute data sensed or generated from their mobile devices, aggregates and fuses the data in the cloud for crowd intelligence extraction and people-centric service delivery". Participants in MCS can either collect "hard" data, stemming from physical internal or external sensors connected to the mobile device, or "soft" data, which refer to human-added information (e.g., annotations, human observations) [22].

We define HQS as a sensing method that collects data or information in high quality trustworthily, which means that the accuracy can be considered error-free and the recording of the data is reliable without failures (like e.g. sensor down times).

MCS can either be used to gather data (or information derived therefrom), that is also professionally measurable by HQS, or aim for gathering data or information that cannot be collected feasibly using professional measurement methods.

3.3 Literature Review on Combining MCS and HQS

We start by presenting the most frequent approach on HS, in which both sensing methods collect the same kind of data and focus on accumulating a large amount of data in order to improve the spatial-temporal coverage and/or data quality. For more clarity, we present them divided further into approaches that add HQ sensors on top of MCS and vice versa, that is adding MCS to an existing HQ sensor network. We address both in the following two paragraphs.

Aiming to overcome limitations of MCS and thus ensuring a stable sensing quality and spatial-temporal coverage, the following approaches add static sensor nodes on top of a MCS system. In order for incentive mechanisms to work in MCS applications, a sufficiently large user base is needed, yet MCS faces the problem that crowd participants do not provide sufficient data at all times (e.g., at night). The hybrid framework (HySense) presented in [23] offers a solution by adding stationary sensor nodes to an environmental monitoring MCS application to ensure spatial-temporal coverage. Users' mobility restrictions can be another source for unreliable sensing quality in MCS, which a HS network, containing both static and uncontrolled mobile nodes, seeks to solve in [24] and [11]. The authors formulate criteria for measuring the sensing service quality in HS, identify relevant influencing factors and develop a theoretical grid-based coverage strategy. In [25] missing sensory data from areas less covered by MCS is also compensated by additional static sensors, which are combined together by means of an interpolation strategy. Evaluation shows that a combining

interpolation with a mix of static and mobile sensors yields better results over a simpler solution where interpolation is based only on data from static sensors.

In the following approaches, the situation is reversed, making MCS the means to improve a static sensor network, aiming to enrich by achieving an improved spatial-temporal coverage with the benefits of low costs and scalability. Four of the identified applications are (prototypically) implemented in an environmental context and one each in a smart city, smart factory and military setting. In [10] the authors introduce a hybrid sensor calibration scheme for MCS applications, to enable more accurate and dense measurements of natural phenomena adding mobile sensing to an existing sensing infrastructure (e.g., weather stations). The proposed scheme was applied to an environmental use case, in which a temperature map of a city was created, resulting in more detailed information than only the infrastructure-based measurements could provide. Another environmental use case (pollution monitoring) is portrayed in [26], in which MCS is presented as an opportunity to ensure better spatial-temporal coverage for stationary sensory networks in a flexible and cost-efficient manner. The authors focus on solving the resulting scheduling problem that faces the challenge of multiple sensor types generating heterogeneous data at different levels of granularity. In order to receive more accurate noise pollution maps with a better spatial-temporal coverage, a middleware solution is introduced in [27], providing a data assimilation technique to estimate noise pollution based on simulation and noise levels measured over both static and uncontrolled mobile sensor, that are added additionally. In [28] the authors examine the potential benefits of combining static and mobile sensors as a participatory sensor network in a use case of measuring the emission of a substance (e.g., pollutant), evaluating their results using mobility models for simulation. Turning away from environmental monitoring towards a smart city context, [12] presents a prototype for enriching stationary infrastructure sensors with smartphone data in order to improve the situation awareness in cities (public safety and sustainability). The authors aim to develop a dynamic sensing platform that intelligently assigns sensing tasks, not covered by static sensors, to smartphone users in a resource-efficient manner. In the context of smart factories a blockchain-based approach for integrating MCS into a static sensing network is introduced in [8], in order to improve the spatial coverage in a scalable and cost-effective manner. The work focuses on resolving the three main challenges, reliability, security and sensory data quality, arising when integrating MCS into a factory. The G-Sense (Global-Sense) architecture [29], prototypically implemented in a military context, integrates mobile sensors into static wireless sensor networks, featuring an algorithm for optimizing the timing for measuring and sending updated data from the mobile device to the server, while meeting the application requirements.

Having presented HS applications aiming to combine MCS and HQS by collecting the some kind of data or information in an accumulating manner, we now list approaches in which both methods are not equally prioritized. They aim to minimize resource input (e.g., sensor deployment, energy consumption) by cost-efficient replacement of the more expensive method, under the condition of a guaranteed minimum sensing quality. In order to improve the sensing quality and eliminate uncertainties resulting from mobility and varying sensing quality of individuals a collaborative sensing approach is presented in [30] and [31] using both mobile phones

and stationary sensors in form of Wireless Sensor Networks (WSN). While [30] introduces an activation scheme for WSN, only enabling stationary sensors when the required sensing quality is not sufficient, [31] focuses on finding optimal locations for wireless sensors in order to minimize the required number of sensors. By solving an optimization problem, the authors in [13] determine the minimal amount of needed additional static sensors and their optimal locations to ensure stable sensing quality and availability, while simultaneously minimize the deployment costs. Trying to overcome limitations of both MCS and static WSN, like network latency, limited lifetime of WSN, costly mobile internet connection, and high battery consumption in the case of MCS, the authors in [32] and [33] introduce a RPL-routing protocol, enabling interaction between MCS and static WSN in a smart city context. In order to optimize activities between data utility (e.g., accuracy) and operational costs (e.g., sensor deployment), a comprehensive planning-based approach with prototype implementation for the combination of mobile devices and in-situ sensors in urban environmental sensing is presented in [34], addressing data generation, upload, and sensor calibration. The authors in [35] introduce a greedy algorithm to solve the dynamic sensor selection problem in a heterogeneous sensor network composed of both mobile sensors and stationary sensors, in terms of location, mobility pattern, energy constraint, and sensing cost.

All approaches listed above use MCS and HQS to obtain the same kind of measurement, for either data replacement or complement. Yet MCS can also be used to generate additional information, which cannot be collected feasibly via HQS methods. Especially all non-physiological measurements fall under additional information (e.g., context information, human perception), as they are difficult or impossible to monitor using traditional sensing networks. A lack of standardisation on data, service and method, uncertainty regarding the measurement of quality, and privacy concerns are common challenges arising when working with "soft" human data [36]. In the following paragraph we present the three applications we have found that use MCS to gain additional information to enrich HQS data.

The integration of spatial-temporal contextual information with human and technical sensor information from a geospatial perspective, which is yet another challenge, is discussed in [9], introducing a model of interactions between humans, the environment and technology in a smart city environment. The MCS application Allergymap [37] is developed in the field of public health monitoring and aims to help people with allergic diseases (e.g., by identification of allergens season, monitoring of treatment process etc.). It combines subjective user input and objective environmental data from fixed stations in a privacy aware manner, outputting a data visualisation in form of a map. The environmental monitoring network introduced in [38] differs from all previously mentioned cases in the fact that it is based on both stationary and mobile high quality sensor nodes. The crowd is used for further data enrichment via a mobile participatory sensing platform, which allows citizens to subjectively report and comment on situations with possible influence on environmental conditions.

The literature review shows, that several approaches on combining MCS and HQS and using its potential already exist, but there is no uniformly structured approach. Most prototypes are implemented in an environmental or smart city context. Our findings

therefore lack in diversity of application, as well as the integration of mobile high-quality sensors.

4 Artifact Description

With the results of our structured literature review at hand, we now continue with the description of our two derived artifacts: The HS classification matrix and the HS framework.

4.1 Hybrid Sensing Classification Matrix

Based on the use cases identified in literature on combing both MCS and HQS we derive a Hybrid Sensing Classification Matrix (Figure 2). We start the development of the HS classification matrix by identifying the features, displayed on the horizontal and vertical axis in Figure 2. For this, we first structure the found literature based on the main aspects of each paper (e.g., minimizing energy consumption, spatial-temporal coverage). In the same process, we extract relevant tasks and requirements for the creation of a first draft of a generic HS framework, which maps the overall process for the value-creating combining of both sensing methods. Based on those two parallel tasks, which are mutually dependent, we inductively derive superordinate distinguishing features for classifying HS systems. While doing so, we have our attention on two aspects. On the one hand, we aim for gaining distinguishing features that help to group the identified approaches into classes, which have similar relevant activities and challenges when concretely applying the HS framework. On the other hand, our features should help to divide the approaches into classes that separate as clearly as possible between approaches, that are most challenging to incorporate into one single framework, due to their varying relevant processing tasks. By inductively testing different distinguishing features we conclude that the overall goal of the data combination (vertical feature: strategic focus) answers the first mentioned aspect well and the used data basis (horizontal feature: data/information) answers the second. Note that other classification features (or more than two) are conceivable, but these have proven to be suitable for the development of the HS framework.

To summarize, while the characteristic depicted on the horizontal axis states whether the data or information collected using both methods is the same or different, the vertical characteristic differentiates between the strategic focus lying on data aggregation or substitution, thus resulting in an equal or unequal prioritization of both methods. This leads to the following four HS classes:

- **Complement:** Both methods collect the same data/information and both methods have equal priority, i.e. the methods do not aim to replace each other, but are combined in order to aggregate a large amount of data to achieve better spatial-temporal coverage, object coverage, or a quality improvement.
- **Direct Replacement:** Both methods collect the same data/information, but not both collection methods are equally prioritized, as they aim to minimize resource input (e.g., stationary sensors, energy consumption) by cost-efficient replacement of the

more expensive method, under the condition of a guaranteed minimum sensing quality.

— **Supplement:** MCS and HQS gather a different kind of data/ information and both methods have equal priority, aiming to enrich each other by adding additional unknown data/ information.

— **Indirect Replacement:** MCS and HQS collect different types of data/information and not both methods are equality prioritized, as they aim for substitution, not aggregation. Although both methods collect a different type of data/information (e.g., sensor data vs. human subjective input), the data/information from one method can be used to approximate the data/information of the other method, thus making it able to replace the measurement.

		Information/ Data	
		Same	**Different**
Strategic Focus	**Aggregation**	**Complement** *Aim: Spatial-temporal coverage, object-based coverage, quality*	**Supplement** *Aim: Additional information (e.g., context, subjective perception)*
	Substitution	**Direct Replacement** *Aim: Cost-efficient resource allocation*	**Indirect Replacement** *Aim: Cost-efficient resource allocation*

Figure 2. Hybrid Sensing Classification Matrix

Note that, whereas the horizontal differentiation between same and different data is clearly assignable, determining the strategical approach is in some cases not equally unambiguous. The overall goal of HS is the combination of both collection methods in a way which maximizes the overall value (which has to be defined individually for each application, as it depends crucially on the strategic focus and could thus target e.g., data quality, coverage or, deployment costs), not replacing one method altogether. Consequently, a complementary component will always play a partial role in the system, even when the strategical approach aims for substitution. This circumstance makes it difficult to assign applications whose combination goal is not clearly communicated in the literature. Due to the fact that HS approaches featured in literature are often not described in full detail, as the research focus may lie on one very specific aspect and not the system as a whole, this is a situation that occurs occasionally. Although this makes a clear assignment difficult in some cases, we have tried to assign the identified approaches consistently to the best of our knowledge, based on the provided information in the respective papers. Table 1 features example HS systems found in literature with type assignment and main target. We found no applications aiming for Indirect Replacement, which leaves its relevance open for discussion, but also possible opportunities open for future research.

**Table 1.** Example HS systems with classification

Ref.	Main aim	Context	Type
[10]	High quality through sensor calibration	Environmental monitoring (temperature map)	Complement
[23]	Spatial-temporal coverage	Environmental monitoring (ozone concentration)	Complement
[26]	Spatial-temporal coverage	Environmental monitoring (pollution mapping)	Complement
[30], [31]	Cost-efficient resource allocation (energy consumption, senor deployment)	Theoretical approach; Evaluation via simulation	Direct Replacement
[34]	Optimization between data utility and operational costs	Community IoT systems in urban sensing	Direct Replacement
[38]	Additional information through subjective user input	Environmental monitoring (air quality)	Supplement
[37]	Additional information through subjective user input for personalized services	Allergy map	Supplement
	Spatial-temporal coverage	RCM (our project)	Complement

4.2 Hybrid Sensing Framework

In the following section, we describe the derivation of the Hybrid Sensing Framework, which aims to generalize the process of combining MCS and HQS in a value-create manner, and present our first draft, illustrated in Figure 3. It visualizes the generalized processing steps in a HS system, including data collection, data combination, data processing, quality evaluation, and quality enhancement tasks for optional system improvement.

As mentioned in the previous section, we build our first draft of the HS framework based on the results of our literature review. For this purpose, we derive essential tasks, requirements and general recurring components from the literature on existing HS applications, yet as stated before the information provided is often incomplete. Therefore we will develop and evaluate an instantiation in a RCM use case (cycle 2), the evaluation results of which we will use for further refining the framework (cycle 3). As we do so, we will continuously check with the literature and the identified HS classes to ensure that all revisions maintain the universal validity of the HS framework.

The final HS framework should include all possible applications combining MCS and HQS, yet up to now, there is a lack of research on the combination of MCS and high-quality data gathered by mobile sensors, making it not possible to extract knowledge on this scenario from literature. We will therefore build a prototype application in our RCM use case to extract further knowledge, which we will use to refine and validate our framework draft in terms of generalization.
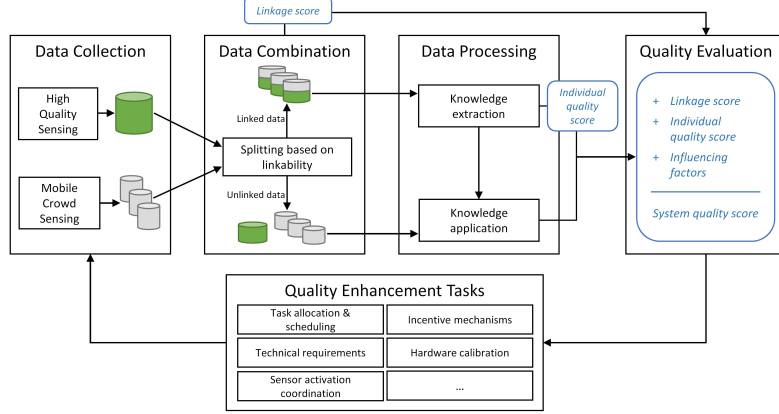
Figure 3. Hybrid Sensing Framework illustrating the processing steps when combining MCS and HQS

Data Collection. MCS gathers human-generated data via mobile devices or connectable sensors [39]. HQS data is collected by either stationary sensor networks (e.g., [10], [11], [23], [24]), mobile high-quality sensors [38] or possibly expert observations. For our RCM application, we developed a crowd app for android phones, making it possible to gather relevant sensor data (e.g., acceleration, gyroscope, GPS) and upload it to our server. The high-quality data is collected, as described in the introduction, by a project partner specialized in RCM, providing us with IRI measurements. We tested both data collection methods extensively, making our operative infrastructure ready to use for large-scale data collection as a basis for the development of data combination methods.

Data Combination. Regardless of whether both methods collect the same or different data, in order to perform a value-creating processing, it must be combined by some kind of link characteristic, which is done in the Data Combination step. In the literature one finds mainly a spatial-temporal link (e.g., in [9], [23], [26]), but it would also be possible to link data, for example, over identical objects featured in images or data connected to a similar situation. In our use case, we have to answer questions on how to define geospatial coverage and how to deal with measurements taken at different times. Aim of this step is to have a data fusion method, merging both data types by geo-coordinates with respect to time and a linkage score (e.g., percentage geospatial coverage within a certain time interval).

Data Processing. From the linked MCS and HQS, data knowledge is extracted (e.g., in the form of a prediction [37], interpolation [25], assimilation [27] or calibration model [10]), which is subsequently used to enhance the unlinked data. When extracting knowledge based on the linked data (which can also be referred to as labeled data) an individual quality score for the specific knowledge extraction is derived, which serves to estimate the quality of the knowledge application on the unlinked data. We will use smartphone data collected via the crowd app, which can be linked to specialist data with respect to space and time, to extract knowledge in form of training an individual IRI

prediction model for every crowd driver. This knowledge, in our case the individual model, will be used to predict the road quality based on data collected by the same driver. Every model has a known accuracy, which is the individual quality score for every driver. Data from crowd workers with no spatial-temporal coverage with high-quality data will be processed with a more generalized model, thus resulting in a greater prediction quality uncertainty.

Quality Evaluation. The creation of suitable and meaningful evaluation scores and identification of influencing factors [11] on the overall system quality, is a core task in the design of a HS system, as they serve as optimization target values when improving the system. We propose to evaluate the quality of the overall system based on the individual quality score, which describes the performance of the Data Processing, and the linkage score, which evaluates the Data Combination, also taking into account potential influencing factors. We will use the linkage score, that has yet to be defined, and the model accuracies as individual quality scores to determine the quality of our RCM system. We will also search for influencing factors in the evaluation.

Quality Enhancement Tasks. Apart from improving the individual quality score (Data Processing) and the linkage score (Data Combination), the overall system quality can also be improved by working on the relevant influencing factors. This can be achieved by means of subtasks which comprise topics like task scheduling and sensor coordination [26], [40], incentive mechanisms and data security [8], or simply stating technical requirements, to name a few. Research on "traditional" standalone MCS already offers extensive research in those domains, yet through changing the initial situation by combining MCS and HQS in a value-creating manner, it will be necessary to revise and extend some already well-researched approaches. Based on the results of our quality evaluation, we will derive and implement possible solutions for quality improvement. If there is too little crowd data available, for example, the development of incentive mechanisms could be of help. If poor prediction model results are based on insufficient linkable data, task allocation mechanisms, which assign crowd workers or specialists to specific sensing tasks, may be beneficial. The quality of smartphone data could be improved by defining a minimal technical standard.

5   Discussion and Outlook

HS gives us new opportunities to fully exploit the possibilities that the spread of mobile devices, and thus MCS, has to offer in a wider field of application. Alongside technical improvements (e.g., sensor quality), the flexibility of approaches is further extended by the possibility of connecting external sensors and integrating smart gadgets. This allows to create new cost-efficient industrial solutions, but also build services for participants, aiming for an improvement in quality of life in general. With the expansion of 5G networks, opportunities also arise for applications based on real-time information and high-speed cloud processing, enabling applications requiring computing power, mobile devices cannot provide (e.g., intelligent hazard detection in traffic based on smartphone image processing). However, involving people into the data collection process not only creates opportunities like improved spatial-temporal coverage, but also raises problems

(e.g., security/privacy issues, data trustworthiness, incentive techniques) [7]. While most resulting challenges have been comprehensively discussed in literature in the domain of MCS, including high-quality sensors and the corresponding effect is so far not dealt with extensively. We now summarize frequently mentioned issues that present challenges for HS applications but also offer opportunities for improved solutions.

— Reliable high-quality data can contribute to enhanced *incentive techniques* by the development of more attractive crowd services. Furthermore, incentives can, when needed, promote the generation of overlapping data by both collection methods.
— Malicious misuse can be detected and prevented more easily, thus improving *data trustworthiness,* through the availability of verified high-quality measurements.
— Subjective human input offers opportunities for various new applications, yet represents a challenge due to the *lack of standardization* when combined with physiological sensor measurements.
— Some HS approaches will require a modified optimal *task allocation* for participants, when including high-quality sensors, and *sensor coordination* strategies, when aiming for a cost-efficient resource allocation.
— Finding suitable *quality and coverage metrics* and relevant influencing factors is crucial and has to be solved individually depending on the field of application, yet the definition of generalized requirements could help practitioners and researches.

We did not find any applications falling into the category of indirect replacement, but we see potential for this HS class when it comes to applications providing public benefit yet facing limited financial resources (e.g., crowd sourcing projects).

6   Conclusion

We conducted a structured literature review on the systematic combination of MCS and HQS (i.e., Hybrid Sensing). This is the first of three cycles in our DS project for creating a generic process for HS systems. We classified existing approaches by their data or information gathered and the strategical approach regarding the sensing method prioritization. This resulted in four main types, differentiating HS approaches: Complement, supplement, direct replacement, and indirect replacement. This categorization contributes a structure to open opportunities and challenges to address in the field of HS. We also presented a HS framework for the structured combination of MCS and HQS and identified the usage of mobile high-quality sensors in HS systems as research gap. Researchers and practitioners may use the framework to structure HS-related projects. The introduction of Quality Enhancement Tasks may be of high relevance for related fields (e.g., Citizen Science) that suffer from data quality issues [41]. We also discussed both contributions in the light of our RCM use case. After having set up the operational infrastructure, we will develop and evaluate an initial combination procedure in cycle two, answering questions regarding temporal-spatial coverage and quality assessment. In the third cycle, we will use the evaluation results to revise the HS framework and improve the RCM application to enable low-cost road maintenance, contributing to road safety by near real time damage detection.

References

1. Roy, S.: How Google Maps is positioned to become the next big content platform, https://www.financialexpress.com/industry/technology/how-google-maps-is-positioned-to-become-the-next-big-content-platform/2087251/ (Accessed: 28.11.2020)
2. Kapadia, A., Kotz, D., Triandopoulos, N.: Opportunistic sensing: Security challenges for the new paradigm. In: First International Conference on Communication Systems and Networks and workshops, pp. 1–10. IEEE (2009)
3. Burke, J.A., Estrin, D., Hansen, M., Parker, A., Nithya, R.A., Reddy, S., Srivastava, M.: Participatory Sensing. Workshop on World-Sensor-Web (WSW): Mobile Device Centric Sensor Networks and Applications (2006)
4. Joly, A., Bonnet, P., Goeau, H., Barbe, J., Selmi, S., Champ, J., Dufour-Kowalski, S., Affouard, A., Carre, J., Molino, J.-F., et al.: A look inside the Pl@ntNet experience. MULTIMEDIA SYSTEMS 22, 751–766 (2016)
5. Pl@ntnet Website, https://plantnet.org/en/ (Accessed: 28.11.2020)
6. Bonnet, P., Joly, A., Faton, J.-M., Brown, S., Kimiti, D., Deneu, B., Servajean, M., Affouard, A., Lombardo, J.-C., Mary, L., et al.: How citizen scientists contribute to monitor protected areas thanks to automatic plant identification tools. Ecological Solutions and Evidence 1 (2020)
7. Shu, L., Chen, Y., Huo, Z., Bergmann, N., Wang, L.: When Mobile Crowd Sensing Meets Traditional Industry. IEEE Access 5, 15300–15307 (2017)
8. Huang, J., Kong, L., Dai, H.-N., Ding, W., Cheng, L., Chen, G., Jin, X., Zeng, P.: Blockchain-Based Mobile Crowd Sensing in Industrial Systems. IEEE Transactions on Industrial Informatics 16, 6553–6563 (2020)
9. Sagl, G., Resch, B., Blaschke, T.: Contextual sensing: Integrating contextual information with human and technical geo-sensor information for smart cities. SENSORS 15, 17013–17035 (2015)
10. Son, S.-C., Lee, B.-T., Ko, S.K., Kang, K.: Hybrid Sensor Calibration Scheme for Mobile Crowdsensing-Based City-Scale Environmental Measurements. ETRI JOURNAL 38, 551–559 (2016)
11. Ding, S., He, X., Wang, J., Qiao, B., Gai, K.: Static node center opportunistic coverage and hexagonal deployment in hybrid crowd sensing. Journal of Signal Processing Systems 86, 251–267 (2017)
12. Liao, C.-C., Hou, T.-F., Lin, T.-Y., Cheng, Y.-J., Erbad, A., Hsu, C.-H., Venkatasubramania, N.: SAIS: Smartphone augmented infrastructure sensing for public safety and sustainability in smart cities. In: Proceedings of the 1st International Workshop on Emerging Multimedia Applications and Services for Smart Cities, pp. 3–8 (2014)
13. Bijarbooneh, F.H., Flener, P., Ngai, E.C.-H., Pearson, J.: An optimisation-based approach for wireless sensor deployment in mobile sensing environments. In: 2012 IEEE Wireless Communications and Networking Conference (WCNC), pp. 2108–2112 (2012)

14. Kuechler, B., Vaishnavi, V.: On theory development in design science research: anatomy of a research project. European Journal of Information Systems 17, 489–504 (2008)
15. Zustandserfassung und -bewertung (ZEB) auf Bundesfernstraßen, https://www.bmvi.de/SharedDocs/DE/Artikel/StB/zustandserfassung-und-bewertung.html (Accessed: 28.11.2020)
16. Klopfenstein, L.C., Delpriori, S., Polidori, P., Sergiacomi, A., Marcozzi, M., Boardman, D., Parfitt, P., Bogliolo, A.: Mobile crowdsensing for road sustainability: exploitability of publicly-sourced data. International Review of Applied Economics 34, 650–671 (2020)
17. Allouch, A., Koubâa, A., Abbes, T., Ammar, A.: Roadsense: Smartphone application to estimate road conditions using accelerometer and gyroscope. IEEE SENSORS JOURNAL 17, 4231–4238 (2017)
18. Forslöf, L., Jones, H.: Roadroid: Continuous road condition monitoring with smart phones. Journal of Civil Engineering and Architecture 9, 485–496 (2015)
19. Webster, J., Watson, R.T.: Analyzing the past to prepare for the future: Writing a literature review. MIS quarterly, xiii–xxiii (2002)
20. vom Brocke, J., Simons, A., Niehaves, B., Niehaves, B., Reimer, K., Plattfaut, R., Cleven, A.: Reconstructing the giant: On the importance of rigour in documenting the literature search process. ECIS 2009 Proceedings (2009)
21. Guo, B., Yu, Z., Zhou, X., Zhang, D.: From participatory sensing to Mobile Crowd Sensing. 2014 IEEE International Conference on Pervasive Computing and Communication Workshops, PERCOM WORKSHOPS 2014 (2014)
22. Rimland, J.C., Hall, D.L., Graham, J.L.: Human cognitive and perceptual factors in JDL level 4 hard / soft data fusion. Proceedings of SPIE - The International Society for Optical Engineering (2012)
23. Han, G., Liu, L., Chan, S., Yu, R., Yang, Y.: HySense: A Hybrid Mobile CrowdSensing Framework for Sensing Opportunities Compensation under Dynamic Coverage Constraint. IEEE Communications Magazine 55, 93–99 (2017)
24. Ding, S., He, X., Wang, J., Dai, W., Wang, X.: Static node center hexagonal deployment in hybrid crowd sensing. 2015 IEEE 17th International Conference on High Performance Computing and Communications, 2015 IEEE 7th International Symposium on Cyberspace Safety and Security and 2015 IEEE 12th International Conference on Embedded Software and Systems, HPCC-CSS-ICESS 2015 (2015)
25. Girolami, M., Chessa, S., Adami, G., Dragone, M., Foschini, L.: Sensing Interpolation Strategies for a Mobile Crowdsensing Platform. In: MobileCloud 2017. 5th IEEE International Conference on Mobile Cloud Computing, Services, and Engineering, pp. 102–108. IEEE (2017)
26. Zhu, Q., Uddin, M.Y.S., Venkatasubramanian, N., Hsu, C.-H.: Spatiotemporal scheduling for crowd augmented urban sensing. In: IEEE INFOCOM 2018-IEEE Conference on Computer Communications, pp. 1997–2005 (2018)
27. Hachem, S., Mallet, V., Ventura, R., Pathak, A., Issarny, V., Raverdy, P.-G., Bhatia, R.: Monitoring Noise Pollution Using the Urban Civics Middleware. In:

BigDataService 2015. IEEE First International Conference on Big Data Computing Service and Applications, pp. 52–61. IEEE (2015)

28. Lent, R., Minero, M., North, R., Barria, J.: Evaluating mobility models in participatory sensing. Proceedings of the Annual International Conference on Mobile Computing and Networking, MOBICOM (2012)

29. Perez, A.J.: An architecture for global ubiquitous sensing (2011)

30. Ngai, E.C.-H., Xiong, J.: Adaptive collaborative sensing using mobile phones and stationary sensors. Proceedings of the International Conference on Dependable Systems and Networks (2011)

31. Ruan, Z., Ngai, E.C.-H., Liu, J.: Wireless sensor network deployment in mobile phones assisted environment. In: 2010 IEEE 18th International Workshop on Quality of Service (IWQoS), pp. 1–9 (2010)

32. Al Sawafi, Y., Touzene, A., Day, K., Alzeidi, N.: Toward hybrid RPL based IoT sensing for smart city. In: 2018 International Conference on Information Networking (ICOIN), pp. 599–604 (2018)

33. Sawafi, Y.A., Touzene, A., Day, K., Alzeidi, N.: Hybrid RPL-based sensing and routing protocol for smart city. International Journal of Pervasive Computing and Communications 16, 279–306 (2020)

34. Zhu, Q.: Exploiting Mobile Plus In-Situ Deployments in Community IoT Systems (2019)

35. Ma, Y., Hou, F., Ma, S., Liu, D.: Dynamic Sensor Selection in Heterogeneous Sensor Network. In: 2016 IEEE 83rd Vehicular Technology Conference (VTC Spring), pp. 1–5 (2016)

36. Resch, B., Blaschke, T.: Fusing human and technical sensor data: Concepts and challenges. SIGSPATIAL Special 7, 29–35 (2015)

37. Kalogiros, L.A., Lagouvardos, K., Nikoletseas, S., Papadopoulos, N., Tzamalis, P.: Allergymap: A Hybrid mHealth Mobile Crowdsensing System for Allergic Diseases Epidemiology : multidisciplinary case study. 2018 IEEE International Conference on Pervasive Computing and Communications Workshops, PerCom Workshops 2018 (2018)

38. Bacco, M., Delmastro, F., Ferro, E., Gotta, A.: Environmental monitoring for smart cities. IEEE SENSORS JOURNAL 17, 7767–7774 (2017)

39. Resch, B.: People as sensors and collective sensing-contextual observations complementing geo-sensor network measurements. In: Progress in location-based services, pp. 391–406. SPRINGER (2013)

40. Ngai, E.C.-H., Xiong, J.: Adaptive collaborative sensing using mobile phones and stationary sensors. In: 2011 IEEE/IFIP 41st International Conference on Dependable Systems and Networks Workshops (DSN-W), pp. 280–285 (2011)

41. Weinhardt, C., Kloker, S., Hinz, O., Aalst, W.M.P.: Citizen Science in Information Systems Research. Business & Information Systems Engineering: The International Journal of WIRTSCHAFTSINFORMATIK 62, 273–277 (2020)