Robotics and Autonomous Systems 90 (2017) 71–85

Contents lists available at ScienceDirect

Robotics and Autonomous Systems

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

A hybrid approach to decision making and information fusion: Combining humans and artificial agents

Frans C.A. Groen [a,*], Gregor Pavlin [b], Andi Winterboer [c], Vanessa Evers [d]

[a] Informatics Institute, University of Amsterdam, Netherlands
[b] D-CIS Lab, Thales Research, Delft, Netherlands
[c] VDI/VDE Innovation + Technik GmbH, Germany
[d] Human Media Interaction, University of Twente, Netherlands

ARTICLE INFO

Article history:
Available online 30 September 2016

Keywords:
Human agent systems
Information fusion
Environmental crisis management
Bayesian Networks
Gas detection

ABSTRACT

This paper argues that hybrid human–agent systems can support powerful solutions to relevant problems such as Environmental Crisis management. However, it shows that such solutions require comprehensive approaches covering different aspects of data processing, model construction and the usage. In particular, the solutions (i) must be able to cope with complex correlations (as different data sources are used) and processing of large amounts of data, (ii) must be robust against modeling imperfections and (iii) human–machine interaction (HMI) approaches must facilitate human use of crisis management tools and reduce the likelihood of miscommunication.

In this paper the relevant problem is an environmental protection application involving the detection and tracking of gases in case of chemical spills in an urban area. We show that a combination of Bayesian Networks, agent paradigm and systematic approaches to implementing HMI, support effective and robust solutions. To better integrate human information and demonstrate the usefulness of user generated crisis response information we developed a social media harvesting interface based on data from Twitter tweets and a visual interface to facilitate human smell classification.

1. Introduction

Social media have changed our world and the way we interact with each other. Despite all the information and communication available in our current society, a close cooperation between humans and intelligent agents is not really present. An Internet of Things makes objects accessible through internet, but does not create a real cooperation. Hybrid human–agent systems, also called Actor Agent communities, can support powerful solutions to real world problems, such as Crisis Management. Such hybrid human–agent systems have to supply, fuse and share information to create a situation assessment enabling reasoning and decision support. The challenge is to design distributed systems where both humans and intelligent agents have access to the same information in a dynamic world model where information is fused and reasoning takes place.

Problems and challenges

In real-world applications of hybrid human–agent systems a number of challenges is present. The world is dynamic and information sources may come and go. So the reasoning and fusion systems should be capable to add and delete components on the fly. Also the internal representation should be capable to cope with time varying information.

Another challenge is that the system, especially in crisis situations, should not be closed but open. Each crisis situation is different with regard to essential and unforeseen information which must be included in the system's world model. The system must have such a structure that this new information can be recorded and tracked in its world model.

The sensory data and human information are distributed. It makes sense to distribute the reasoning system as well, so that there is a graceful degradation of the system and no single point of failure, due to miscommunication or disruption of the system. Often multidisciplinary knowledge is required to develop the components of such a distributed system. This collaborative design however, adds to the complexity of the system.

The interaction with humans is another important aspect. Human information is fused with sensory agent information. The question is not only how to merge the information, but also how to interact optimally for instance through a smart mobile phone application. The distributed reasoning system and the interaction design can have a general setup. The interaction and perception

* Corresponding author.
E-mail address: fcagroen@gmail.com (F.C.A. Groen).

http://dx.doi.org/10.1016/j.robot.2016.08.009
0921-8890/

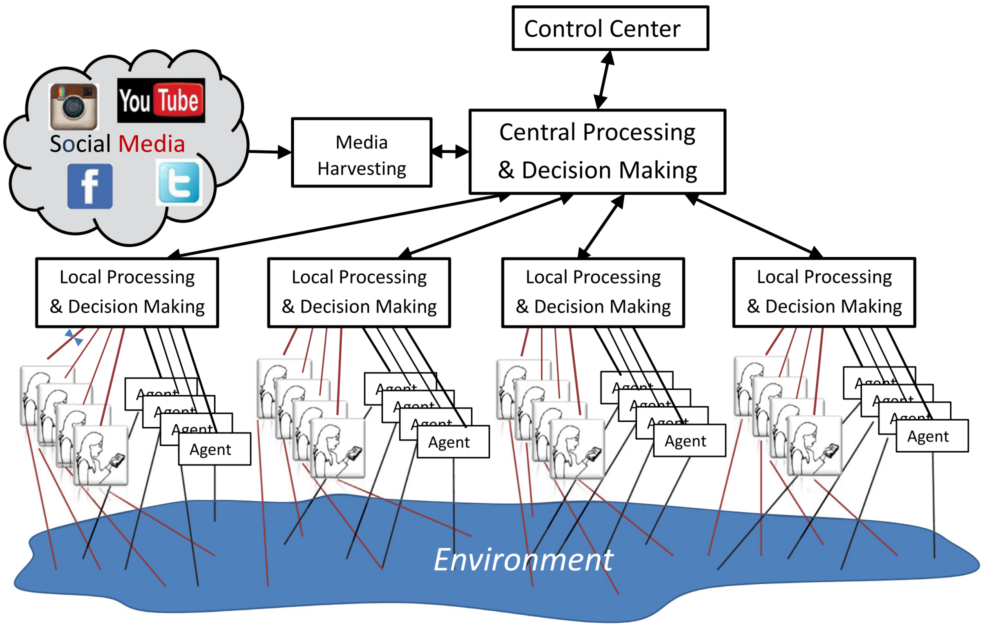


Fig. 1. General architecture of hybrid human–agent systems.

are dependent on the modalities that the application/platform provides but the interaction should be designed to deal with this.

Architecture

In Fig. 1 the general architecture of a distributed human–agent system is given. Both humans and agents observe the environment and can interact with it. The information from both the humans and the agents can be locally processed and decisions can be taken, which leads to directions for actions or requests for information. The advantage of introducing a local processing step is a graceful degradation of the system when communication with a central processing system becomes cumbersome. Social media nowadays are also an important source of information. Information harvested from social media can be fused centrally with the locally processed information and presented to a control and command center. From the command center instructions can send back again to people through their devices and social media.

Application

In this paper we will focus on the domain of environmental protection and a human–agent system for monitoring air pollution (Winterboer et al. [1]). In the Netherlands, as well as in other densely populated countries, locations of harbors and industries processing chemicals are often in close proximity to places where people live. In the case of a serious chemical spill, a system for the detection and tracking of potentially dangerous gases is essential. People in critical regions can be informed through their mobile phone and safety instructions can be issued. Information to detect and track the gases can be obtained from different types of gas sensors present at various locations. Decisions have to be made about the location of the gas, the type of gas and what measures need to be taken. However, the development of such a system is also highly dependent on reliable smell descriptions from people in the vicinity.

There are no devices that would reliably detect and identify different types of gases. State of the art gas sensors are usually noisy and they can fail due to different reasons, such as physical damage or unusual operating conditions. As the currently available sensor networks are typically not dense enough to support robust detection and tracking, situation assessment is often strongly supported by information obtained from humans, who can be considered as omnipresent information sources that can provide rich and useful information. By exploiting the existing communication technology and social media, we can access humans in the affected area who can provide valuable information on the situation.

This paper addresses the problem of detecting and localizing the pollution source, without estimating the extent of the plume. In particular, the challenge is to localize the source without concentration measurements. This is quite a unique approach as most of the state of the art localization methods rely on calibrated sensors, which is very expensive and often impractical in the targeted domains. Complementary work on the modeling of gas concentrations and estimation of the gas plume extent can be found in Asadi et al. [2] and Lilienthal et al. [3].

Different types of heterogeneous information, delivered by sensors and human beings have to be merged. Bayesian Networks can be used for that. In a Bayesian network the relationship is represented between the involved variables. They also form the statistical framework to determine the probabilities of the values of the variables and to fuse information.

2. Problem description

In this paper the challenges and various methods will be illustrated with the help of an environmental crisis management use case in which the presence of a gas plume must be detected and the pollution source localized. This must be achieved by (i) using a network of cheap chemical sensors, which are not calibrated and (ii) human reports about various olfactory observations collected through crowd sourcing. In combination with suitable algorithms, such sources can be used for simple detection of anomalies, i.e. unusual gas mixtures caused by industrial pollution. As the system receives a trigger, a positive detection of a specific gas or an industrial pollution, potential pollution sources are determined by taking into account the average wind speed, wind direction and simple gas dispersion models. So the goal of the overall hybrid system is to collect relevant data in form of a set of observations E and infer the hidden causes h_k of these observations. For each possible source s_k we create a hypothesis h_k that an invisible gas of an industrial origin is released at a source s_k. Moreover, as we are dealing with noisy sources and uncertain domain knowledge,

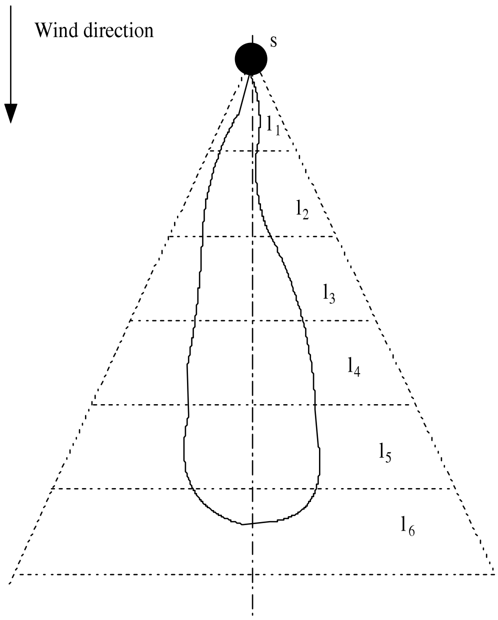


Fig. 2. A simple model of the area influenced by a plume originating at source s. The model is discretized into multiple segments denoted by l_i.

we compute posterior probability $P(h_k|E_k)$ stating that h_k is true given the entire relevant set of evidence E_k. $P(h_k|E_k)$ can be used in different ways in a decision making process, dependent on the domain/principles used for making decisions. For the sake of simplicity we assume that when probability $P(h_k|E_k)$ exceeds a certain threshold we assume that h_k represents the cause of our observations.[1] The computation of $P(h_k|E_k)$ is based on various assumptions about the gas propagation and the operation of the sensors and human observers [4,5] and wind direction and wind speed sensors.

For each hypothesis h_k we create a triangular area whose vertex is defined through the location of the hypothetical source s_k and its axis is parallel to the wind (see example in Fig. 2). The area is further subdivided into segments denoted by $l_1, \ldots, l_n$. An observation is associated with a specific hypothesis h_k if it is contained in any of the segments of the corresponding downwind region R_k. E_k denotes all observations that can be associated with h_k in this way.

If the triggering source (sensor or a human report) is located within the cone associated with a potential source s_k, the system creates a hypothesis h_k.

Note that the cone is an extremely crude gas dispersion model which represents an area in which gas concentrations are likely to exceed the levels required for the detection of gases/anomalies. The model is chosen such that it covers an area which is significantly greater than the actual area in which such concentrations would be observed. The justification for this simplifying assumption is provided in [5] while [4] presents experimental results which confirm that such simplifications do not jeopardize the overall detection/localization accuracy in the targeted domains.

The computation of $P(h_k|E_k)$ is based on a causal Bayesian domain model which describes gas propagation and typical observation sequences throughout the segments of the corresponding region R_k. (i.e. patterns), given hypothesis h_k were true. Bayesian networks (Pearl [6]) are well suited for that purpose. There are a number of advantages of using Bayesian Networks (BN) for reasoning about the hidden causes of observations and the fusion of different sources of information. Namely, often the data can be viewed as outcomes of causal stochastic processes where the causal relations can be described through conditional probability distributions and Directed Acyclic Graphs (DAG). BNs can be obtained through expert introspection or machine learning. The interpretation of the networks by a human observer is easy. This holds both for the variables as well as for the structure revealing the dependences between the variables. Fusion can be done at different levels and with incomplete information. BNs can relate very heterogeneous variables and realize efficient inference with uncertain relations. Bayesian Networks are defined through their structure: a Directed Acyclic Graphs (DAG) and Conditional Probability Tables (CPT).

[1] Alternatively, $P(h_k|E_k)$ can be used for the ranking of hypotheses. See a more elaborate discussion on this in [4].

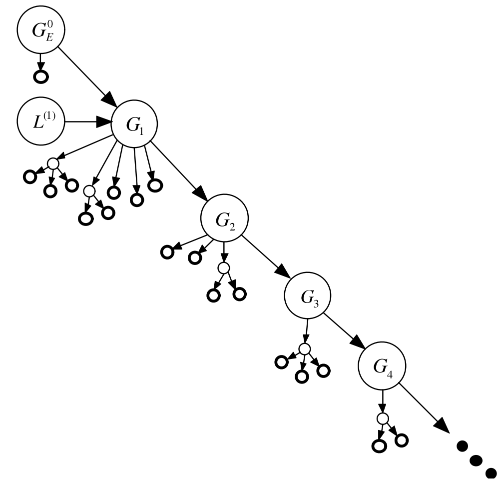


Fig. 3. A simplified domain model describing gas propagation through a discretized area and capturing relations between the hidden states and the observations.

Contrary to common approaches to source localization (e.g. [7–9]), the chosen solution does not rely on concentration/intensity measurements since the installation of sufficiently large and dense networks of calibrated chemical sensors is economically infeasible in the targeted domains. Therefore, the presented approach supports detection and source localization based on complex patterns of heterogeneous binary observations, such as outputs of different types of chemical detectors based on simple uncalibrated chemical sensors and human reports.

3. Bayesian models and inference

In the presented use case, the models are based on a crude discretization of the regions R_k influenced by the gas plume, i.e. regions within which critical gas concentrations are exceeded (see Fig. 2). In addition, temporal aspects of the dynamic processes are captured with the help of time slices, each representing a snapshot of the states at a specific time step. This discretization is captured by the causal model shown in Fig. 3.

The model can be interpreted as follows:

- Each node G_i represents the presence of the gas at location l_i, i.e. segment in the discretized area shown in Fig. 2. More precisely, $G_i = true$ if the plume front has reached the boundary between segments l_i and l_{i+1}. The node remains in the same state until the end of the estimation process.
- Temporal aspects of the gas propagation are not explicitly represented by the model. Instead, the gas propagation is captured through a sequence of segment activations. The

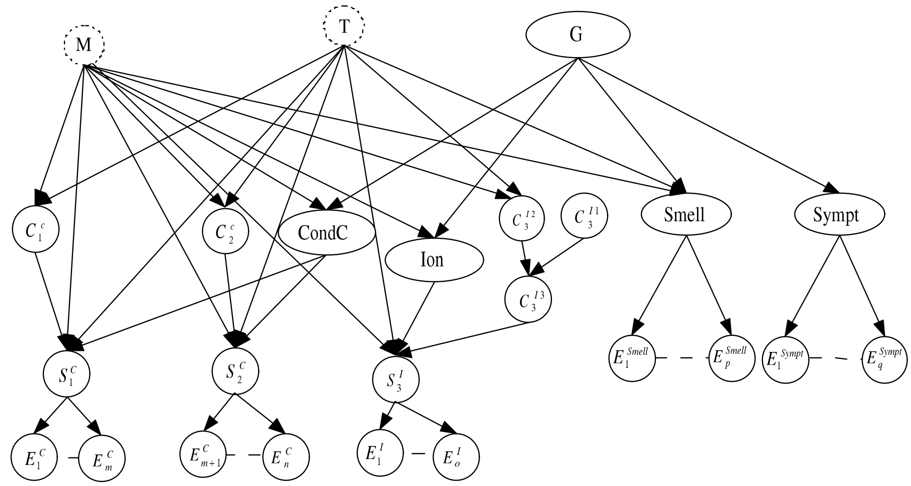


Fig. 4. A composite sensor model capturing the observations at a specific location.

evidence at segment l_i is not processed and fused with the overall model if it is obtained prior to the time at which the hypothetical plume front would reach l_i and l_{i+1}. This estimated time can be based on appropriate gas propagation model. After this point in time, all the evidence collected in segment l_i is treated as if it belonged to a single time slice, which starts at the time of the segment activation. This is a consequence of the fact that in the case of a plume the gas would remain present at each segment until the end of the process.

The causal model in Fig. 3 also specifies relations between each G_i and the observations collected at the corresponding location l_i. Little bold circles in Fig. 3 represent the observations. Each small network attached to node G_i captures the causal stochastic process producing observations at location l_i. Note that observation models can be arbitrarily complex causal models themselves, each capturing correlations between different types of observations obtained at a particular location. An example of such a complex observation model is shown in Fig. 4, which captures the correlations between observations of different types, obtained by sensors using conductivity effect of a semiconductor elements, ionization sensors as well as human reports of smells and health symptoms. The derivation of the above mentioned models is thoroughly discussed in [5,10].

For the sake of brevity, the discussion in this paper focuses on the composite sensor models, as the one shown in Fig. 4, each associated with a specific segment l_i. Monitoring at each segment l_i can be viewed as a causal stochastic process, where hidden events cause observations according to certain probability distributions. Let us assume that the presence of a certain toxic gas causes conditions under which a certain type of semiconductors results in a distinctive conductivity. Similarly, a particular conductivity could be observed in an ionized gas mixture if the toxic gas is present. In this paper we assume two types of sensors, evaluating conductivity in semiconductors and in an ionized gas mixture, respectively. Introduction of a sensor measuring a particular type of conductivity will spawn various processes in the sensor's electronic circuitry which in turn will result in a certain state of the sensor. Dependent on the sensor state we will obtain a sequence of reports, either confirming or refuting the presence of the gas. Similarly, humans are likely to report about typical symptoms (e.g. smell). Such a causal process can be described through the graph shown in Fig. 4, where each node represents a binary variable; e.g. $G = true$ if the gas is present, otherwise $G = false$. The situation under which a semiconductor element and ionized gas mixture feature typical conductivity is represented by variables *CondC* and *Ion*, respectively. States of the binary variable *CondC* correspond to the situations where electrical current under ideal circumstances would either exceed some detection threshold (i.e. $CondC = true$) or remain below that threshold (i.e. $CondC = false$). The states of the ith sensor of type x are represented by S_i^x while a sequence of sensory reports is denoted by binary variables $E_1^x, \ldots, E_n^x$; $E_k^x = true$ if a report confirms the presence of the Gas. In this example, subgraphs containing nodes $S_1^C, C_1^C, E_1^C, \ldots, E_m^C$ and $S_2^C, C_2^C, E_{m+1}^C, \ldots, E_n^C$ describe processes in two sensors measuring the conductivity of local semiconductor elements. Subgraph consisting of nodes $S_3^I, C_{13}^I, C_{23}^I, C_{33}^I, E_1^I, \ldots, E_o^I$, on the other hand, corresponds to the third sensor measuring conductivity of the ionized gas mixture. Variables S_1^C and S_2^C denote the measured conductivity on the semiconductor elements in the first two sensors while S_3^I denotes the measured conductivity of the ionized gas mixture in the third sensor. Moreover, variables $C_1^C, C_2^C, C_{13}^I, C_{23}^I, C_{33}^I$ represent the states of critical electronic components of the three sensors. We also assume that the causal process is influenced by the air humidity and temperature represented by variables M and T, respectively. Note that with each sensor an independent local causal process is introduced to the domain. Besides the sensors we assume that there are humans in the area, who have olfactory reactions to G and who submit reports of what they smell via a call service, app or a web-interface. The states of binary variable *Smell* represent situations in which people familiar with a typical smell of G either do or do not recognize the smell. Moreover, each individual report is represented by a node E_i^{smell}. Similarly, first aid workers might be able to report about health symptoms, which are typical results of exposure to G. A situation in which observable symptoms take place is denoted by variable *Sympt* and the reports are denoted by variables E_i^{Sympt}.

Bayesian Networks can cope with uncertainties in modeling (i.e. the domain knowledge) and observations. However, each source must be explicitly captured in a BN, which is challenging in domains where information sources are dynamic (sensors are added at runtime). So for each constellation of information sources we need a specific BN. In addition, large quantities of heterogeneous information accessed through the existing communication and sensing infrastructure often require large BNs which in turn require significant processing and communication resources. Both, the dynamics of the environment and the processing complexity can be tackled by introducing a modular approach to modeling and processing.

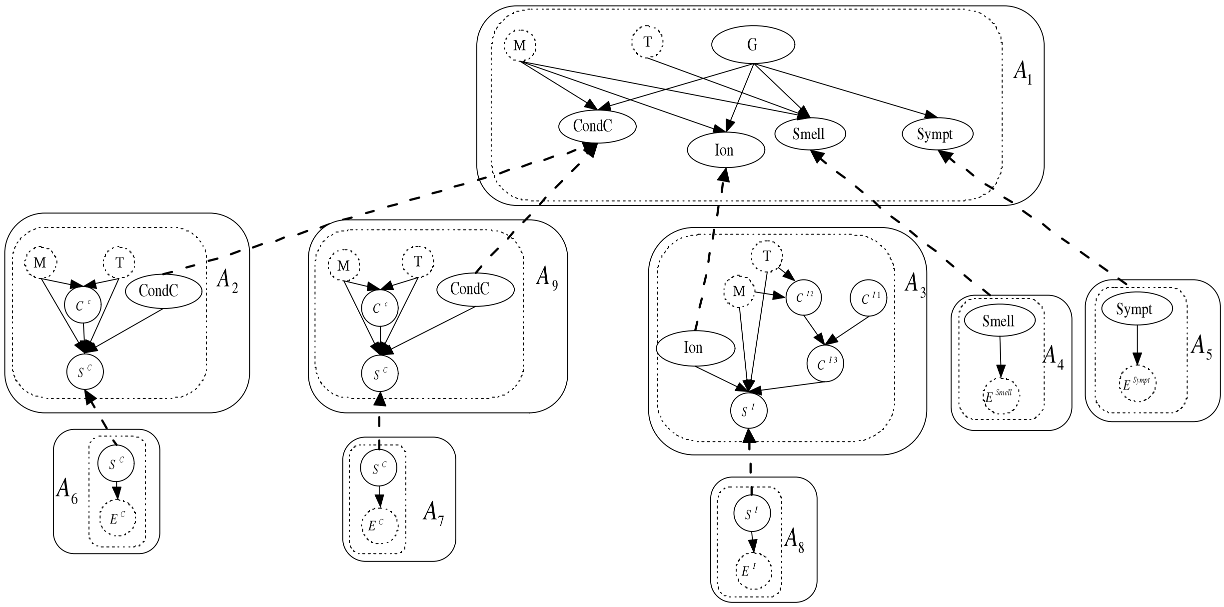


Fig. 5. An organization of fusion agents implementing a distributed causal model. Each solid rectangle corresponds to an agent with a local modeling fragment represented by dashed rectangles. Dashed arrows show the flow of the inter-agent fusion messages.

4. Distributed fusion

The challenges introduced by large quantities of data stemming from dynamic sources can be efficiently tackled by creating complex domain models through a combination of simpler fusion modules at runtime, as new sources become available. The resulting system of loosely coupled modules fuses data obtained from disparate modules, such that each new piece of evidence is correctly reflected in the probability distributions over the hypotheses of interest. Such data-driven fusion is based on information flows between disparate fusion modules that exchange messages carrying probability distributions over variables representing correlated phenomena in the domain. Such modularization requires (i) a sound decomposition of BN models and inference processes and (ii) a platform that supports non-trivial information flows and self-organization.

In this paper we assume the modularization approach introduced in [10] that allows correct inference in a system of loosely coupled BN models (see for example Fig. 5). The modularization is based on simple design rules based on the theory on factor graphs [11]. The resulting system of BN modules supports exact globally coherent Bayesian inference without compilation of secondary inference structures spanning multiple processing modules, such as Junction Trees [12,13]. Xiang [13] introduces a method for inference in multiply sectioned Bayesian Networks and Paskin and Guestrin [12] proposed a runtime compilation of junction trees. These approaches require compilation of secondary inference structures, which can be computationally expensive and time consuming. In addition, the approach from [12] requires prior knowledge of all information sources. Consequently, these approaches do not support quick adaptation of fusion systems and cannot efficiently cope with domains where information source constellations can change at runtime. In [14] we propose a method to reduce the dependences between the modules through instantiation of variables. Markov Boundaries provide the means for systematic reduction of dependences between local BNs. The approach is related to loop cut-set conditioning [6]. Contrary to the other loop cut-set conditioning methods, we assume that certain variables in inference modules are instantiated by using hard evidence, thus avoiding combinatorial explosion which is typical for the loop cut-set method in the case of many loops in the networks. In the resulting systems, each module executes inference on a local BN by using any standard inference method, including local junction trees, which can be compiled prior to the operation. The modules achieve globally coherent inference by exchanging marginal probabilities based on local inference [10,15].

Complementary fusion modules autonomously form meaningful distributed fusion systems. Fusion in such an assembled network can easily be distributed throughout several machines thus avoiding processing and communication bottlenecks. Beside sound and efficient fusion algorithms, basic fusion modules must support also efficient communication and cooperation protocols. In addition, a distributed fusion system should be able to adapt to the current situation autonomously. Therefore, modules should form fusion systems consisting of relevant modules autonomously and they should be able to reason about resource allocation with respect to sensing and processing capacity. In order to be able to cope with such complex functionality in a systematic way, we make use of the multi agent systems paradigm (Jennings et al. [16]). Each fusion module is an autonomous agent that provides a certain fusion service and has the logic to form meaningful fusion systems through service composition.

- A **fusion agent** A_i *is a processing unit, a module, which can compute probability distributions over variables* V_i *in its local BN.*
- *Each agent* A_i *maintains a set of* ***service variables*** $R_i \subset V_i$ *and a set of* ***input variables*** $L_i \subset V_i$.
- *Each agent* A_i *can compute marginal posterior probabilities over the local service variables.*
- *Two agents can exchange their local estimates of marginal (posterior) distributions for any variable contained in the local BNs of both agents.*

In Fig. 5 we can identify nine agents. Agent A_1 has service variable $R_1 = G$ and the following input variables $L_1 = \{CondC, Ion, Smell, Sympt\}$. Variables in L_1 on the other hand, correspond to service variables (i.e. outputs) of agents A_2, A_3, A_4, A_5, A_9, i.e. $R_2 = R_9 = CondC$, $R_3 = Ion$, $R_4 = Smell$, $R_5 = Sympt$. The agents exchange messages carrying locally estimated probability distributions over these variables.

Agents wrap information sources and provide uniform communication and fusion protocols. As new information sources wrapped by agents enter the scene, the domain models of fusion systems are adapted on the fly, without any centralized control.

A. Fusion Organization

Each fusion process depends on the constellation of cooperating agents, which corresponds to a particular problem or task decomposition. In this context we use the concept of a *Fusion Organization*. A particular organization Ω is a function of a given query or hypothesis E and a set of available agents (Pavlin et al. [14]).

A global task of a fusion organization Ω is computation of the probability distribution $P(H|E)$ over some hypothesis variable $H \in R_i$ of agent A_i which correctly reflects the entire evidence set E. This is only the case if $P(H|E) = P'(H|E)$ where $P'(H|E)$ is computed through propagation of the entire evidence E in a monolithic BN which correctly captures all dependences and independences that exist between variables in and between the agent's local models in Ω. This requires well-defined cooperation of fusion agents since evidence E corresponds to instantiations of variables in different agents in a fusion organization Ω.

In the presented use case the BN modules are used for the creation of composite observation models as well as for the fusion using models of dynamic processes. For each segment l_i of the downwind area A_h a set of BN modules is assigned. Such a set of modules supports inference about the presence of a gas at l_i by considering specific data sources, such as sensors of different types and reports from humans associated with this location. Fig. 5 shows an example of such a set of modules collaboratively estimating the presence of a toxic gas at a specific location by exchanging partial fusion results. In fact, this system of fusion modules supports Bayesian inference which is equivalent to inference based on the monolithic observation model from Fig. 4. Note also that each composite observation model is an organization of collaborating modules formed at runtime through service discovery as the relevant data sources become available. Moreover, the results of such composite observation models for different locations $l_1, \ldots, l_n$ within the hypothetical plume area A_h are fed to module D_h that supports inference about the dynamic gas propagation processes. Fig. 6 shows a set of modules implementing a fusion system dedicated to a specific hypothesis h associated with a specific downwind area A_h (see example in Fig. 2). In such a system, module D_h estimates the likelihood of a leak by exploiting the knowledge of gas propagation captured by the hidden variables of the model shown in Fig. 3. D_h fuses beliefs of fusion modules dedicated to different segments (i.e. locations) in A_h, which estimate the likelihood of the presence of a specific gas at each location l_i by fusing all available observations at l_i.

Importance of variables in a BN

The acquisition of some variables can be costly, time consuming or involve risks. So it could be very helpful to know the maximum impact that one variable could have on another variable in the network. It also gives the possibility to prioritize missing evidence (relevant to a decision). This is for instance used in the visual human smell interface in the next section. It can also speed up the inference by a further reduction of the complexity of the network (Engelen [17]). A practical problem is that the computation of the maximum impact of one variable on another given certain evidence is infeasible for larger BNs. In Gosliga et al. [18] an approach is proposed for a fast approximation method to determine the maximum impact that one variable can have on another. This efficient approximation never underestimates the impact.

Cost of communication

Particularly in the early stages of a crisis, communication channels may be damaged. Communication channels may also suffer from limited bandwidth and delays. In dynamic situations the relevance of information will also decrease when time passes. So in such cases there will be a tradeoff between the cost of communication and the relevance of the information send. Agents should be aware of the usefulness of sending the information given the impact of variables and the delay (Foeken et al. [19]).

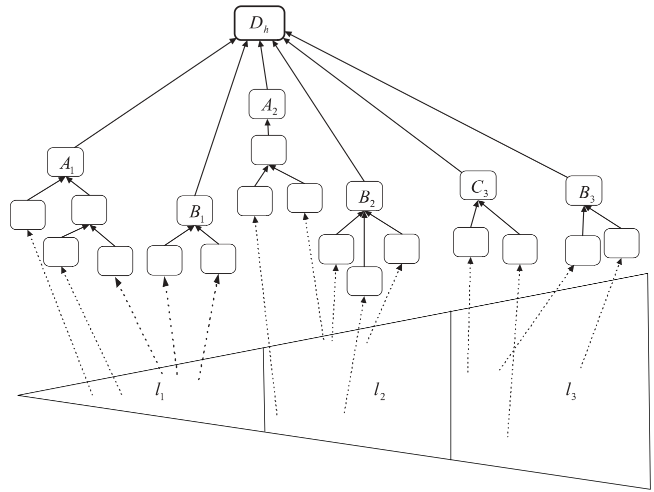


Fig. 6. An organization of fusion modules implementing a distributed causal model of a stochastic process generating observations at different locations $l_1, \ldots, l_n$ and times within a hypothetical area Ai. Note, each box corresponds to an arbitrarily complex Bayesian Network fragment.

5. Modeling and robust information fusion

It turns out that, by explicitly considering causality, the critical dependences in monitoring processes can easily be determined. In such domains causal processes are well understood, since they are to a great extent created through designers of monitoring systems and operators; such processes are created by (i) putting together various man-made components, such as sensors, communication systems and (ii) exploitation of well known procedures, such as asking people that call an emergency center specific questions about the symptoms and smells and other observations. We can safely assume that sensor developers understand main aspects of causal mechanisms in the sensing devices while the professionals in an estimation process follow well established procedures. Model construction consists of (i) the development of causal graphs, which capture direct dependences between the variables, i.e. qualitative knowledge, and (ii) the determination of the parameters which capture the strength of the causal (stochastic) influences, i.e. quantitative knowledge.

Construction of robust causal domain models

Construction of causal domain models for gas detection is exploiting the fact that with each sensor or a human reporter a new local causal process is introduced to the domain. By introducing a new gas sensor, the gas initiates processes on the measuring element and in the sensor's circuitry which eventually produces sensor reports. For example, the process introduced by the third sensor is captured by the graph fragment consisting of variables $S_3^{Ion}, C_3^{Ion}, E_{n+1}^{Ion}, \ldots, E_o^{Ion}$ shown in Fig. 4. In principle, the local processes within a sensor become part of the overall distributed mechanism generating relevant observations. Similar reasoning applies to human observers. Moreover, each type of sensors usually provides observations about a single phenomenon represented by a process variable, such as, for example, Cond and Ion in Fig. 4. Thus several sensors of the same type are influenced by a single phenomenon resulting from a causal process and by a few boundary

Table 1
CPT capturing the estimated perception model for human observers.

$P(E_i^{Smell} \mid G)$	$G = true$	$G = false$
$E_i^{Smell} = yes$	0.75	0.34
$E_i^{Smell} = no$	0.25	0.66

conditions represented by context variables, such as M and T in Fig. 4. In addition, components and reports of one sensor do not influence components and reports of another sensor. Thus, we can represent a process generating observations of a certain type with a network fragment which is sparsely connected to other fragments. Only variables for which direct dependences exist are connected directly. In other words, the resulting causal graph describing the overall observation generating process is sparsely connected.

The modeling parameters, i.e. the conditional probabilities defining the CPTs, could be obtained with the help of the Maximum likelihood estimation method [20] or through expert introspection. The former is a sound solution, however, it is viable if all variables in the model can be observed. In case of partially observable models EM algorithm can be used [21]. Sometimes, experts can estimate the conditional probabilities. In principle, the CPTs relating variables corresponding to the sensor components, observed phenomena and the observations can be estimated by repeated experiments. In some cases, this approach can be also practical for the estimation of CPTs relating human reports and the presence of gases. For example, we could extract such relations from the database compiled by the DCMR milieudienst in Rijnmond, an environmental protection agency in the Port of Rotterdam. The database captures complaints/reports of citizens collected during incidents with known causes. The database allowed estimation of conditional probability distributions for simple reports from citizens indicating the presence or absence of an industrial air pollution, i.e. an anomaly caused by an abnormal concentration of any substance that can be classified as gas, chemical vapor or oil derivatives, which are typically released by the industry. In this case GasX = true corresponds to the presence of such an anomaly. In principle we counted how often complainers responded with yes/no to a question "Does the smell remind you of a chemical or oil or gas?". This question was asked via a web-page or by an automated response system when complainers call a special number provided by the DCMR. In the latter case the complainers could respond by pressing various options, such as 1 for yes, 2 for no and 3 for do not know. By using the DCMR database we extracted the perception model $P(E_i^{Smell}|G)$ for citizens shown in Table 1. The estimation of the parameters in Table 1 was based on 586 incidents for which the cause of complaints was known. The 95% confidence intervals for the parameters were ± 0.03 for $P(E_i^{Smell}|G = true)$ and ± 0.037 for $P(E_i^{Smell}|G = false)$, respectively.

If parameters in a BN are identical to the true distributions over the modeled variables, then a Bayesian classifier is optimal (Duda & Hart [20]). Unfortunately, in many real world applications, it is very difficult or even impossible to obtain probabilistic models that precisely capture the true probability distribution over the phenomena in the observed domains. Training data sets are finite and human experts cannot precisely specify the domain models.

However, by considering the theory of BNs it was shown that reliable classification can be achieved even if the parameters significantly deviate from the true probabilities, as long as the used BNs satisfy simple conditions [14]. *Key to robust inference are relations between the true conditional probability distributions and the distributions captured by the used CPTs* [14]. Let us assume a true conditional probability distribution $P(E|C)$ between variables C and E whose states c_i and e_j correspond to causes and effects, respectively. It can be shown that a system is robust if the used CPTs $P(E|C)$ and the true distributions $\hat{P}(E|C)$ satisfy simple conditions:

$$\forall e_j \in E : \text{argmax}_{c_i} P(e_j|c_i) = \text{argmax}_{c_i} \hat{P}(e_j|c_i) \quad (1)$$

and

$$0.5 < \sum_{e_j \in B_{c_i}} P(e_j|c_i), \quad (2)$$

where B_{ci} denotes the set of all states of E for which the likelihood of state c_i is maximum: $B_{ci} = \{e_k | \forall c_i \neq c_j : \hat{P}(e_k|c_i) > \hat{P}(e_k|c_j)\}$. In case of binary variables the relations (1) and (2) are satisfied if, in both, the CPT describing the true conditional probabilities and the CPT from the model, the elements of the same diagonal exceed 0.5. For example, let us assume a true distribution $\hat{P}(E|C)$ over binary variables E and C: $\hat{P}(e_1|c_1) = 0.7$, $\hat{P}(e_2|c_1) = 0.3$, $\hat{P}(e_1|c_2) = 0.4$ and $\hat{P}(e_2|c_2) = 0.6$. We say that a CPT $\hat{P}(E|C)$ in a **model** correctly captures relations between the true probabilities if its parameters satisfy $\hat{P}(e_1|c_1) > 0.5$ *and* $\hat{P}(e_2|c_2) > 0.5$. In the targeted domains it is often plausible to assume that relations (1) and (2) can reliably be identified by experts or extracted from relatively small data sets with the help of machine learning techniques. This is for example the case with the CPT shown in Table 1. The relations stay the same for any combination of parameters in the 95% confidence interval. Moreover, in [14] it was shown that if the BN corresponds to a *factor tree* whose root is the hypothesis variable, then the expected classification accuracy asymptotically approaches 1 with the growing number of branches rooted in the hypothesis variable if relations (1) and (2) are satisfied for all CPTs in the BN. This is a consequence of the inherent properties of the BNs [6]. A factorization represented by a BN corresponds to a factor tree if the BN has a DAG with a tree topology or appropriate variables in multiply connected BNs are instantiated. Example of the latter is instantiation of variables M and T in the multiply connected model from Fig. 5. If many information sources are available, we can obtain BNs corresponding to factor trees with large branching factors which makes fusion reliable even if we use CPT parameters that deviate from the true distributions significantly.

The robustness can be illustrated with a simple example, where the detection of a gas in a certain area is based on complaints only and the true distribution $\hat{P}(E_i^{Smell}|G)$ is given by Table 1 and the associated 95% confidence intervals.[2] Thus, we would instantiate only the leave nodes corresponding to smell observations in the model from Fig. 5. This would be equivalent to reasoning with a naive BN. If 0.5 were used as the decision threshold, the detection would be equivalent to simple majority voting [14]. Consequently, we can show that the lower bound on the detection performance would asymptotically approach 1 for arbitrary CPTs $P(E_i^{Smell}|G)$ as long as they satisfy the following relations: $P(E_i^{Smell} = true|G = true) = P(E_i^{Smell} = false|G = false) > 0.5$, the same relations that can be found between the true conditional probabilities from Table 1. For 10 reports the expected detection accuracy would exceed 0.88 while the accuracy would exceed 0.99 if we obtained more than 30 reports.

Evaluation of the system performance

The detection and localization accuracy of the presented decentralized system depend on multiple factors, such as the distribution of data sources within the area that is considered in the estimation, the quality of data sources, modeling inaccuracies as well as the criteria used for decision making. Due to usual heterogeneity of the data and a large number of possible situations the system can be exposed to, the evaluation of the overall performance requires huge numbers of experiments (large numbers of permutations of possible constellations of data sources). Despite such complexity,

2 In this experiment we assume that the true distribution is defined by the worst case parameters defined in the 95% confidence intervals of the parameters in Table 1. In this case, the probability of a correct answer given the presence of the gas would be $0.75 - 0.03 = 0.72$. This number defines the chance of having a report correctly indicating the presence of the gas.

the evaluation of the expected system performance can be broken down into smaller problems. This is the case if the processes can be described with BNs that feature relatively sparse graphs. In the running use case, passive gas detectors do not influence each other[3]; i.e. they are conditionally independent given the hidden variables representing the gas propagation. Because of such loose coupling of the phenomena in data generating processes, the overall models used for the fusion in the presented example can be viewed as a composition of (i) loosely coupled components that describe the detection processes and (ii) components that describe the dynamic gas propagation processes. Examples of the former are shown in Fig. 5 while an example of a dynamic model is captured by the variables G_i and CPTs $P(G_i|G_{i-1})$ in the model shown in Fig. 3. The inference about dynamic gas propagation processes and the leak localization is carried out by module D_h shown in Fig. 6.

The expected accuracy of individual detectors and their components can be evaluated in controlled experiments in labs or outdoors. For each detector type this can be captured by the causal models in a statistically sound manner by using Maximum Likelihood Estimation methods. Moreover, as it was illustrated in the previous section, effective observation models can be obtained also for humans reporting on their perception on chemical substances.

However, the estimation of the expected accuracy of the modules estimating the states in dynamic processes is more challenging as such modules interpret spatio-temporal data patterns. For example module D_h in Fig. 6 relies on data originating from many distributed detectors whose outputs depend on the relative position w.r.t. to the leak and the weather conditions. In addition, the performance of this module depends on the efficiency of the data association processes. In order to carry out statistically meaningful evaluation the module has to be exposed to a large number of possible situations.

As it is impractical and often impossible to obtain sufficient quantities of real world data with the ground truth, the estimation of the expected performance of the overall system is carried out through a large number of simulations. Many different situations are simulated through sampling of data based on a systematic variation of the various modeling parameters. In this way we can generate many data sets for known ground truth and estimate the detection and localization accuracy. Clearly, the simulations cannot perfectly capture all conditions in the real world. Instead, they are based on many simplifying assumptions. So the question is whether the results based on simulations allow any reliable conclusions about the expected performance in the real world applications. This is possible if the following is the case:

- Sensor models for the used detectors are based on real world data stemming from tests under different conditions in controlled experiments.
- It can be shown that the expected performance is insensitive to various assumptions about the physical world. This is done by using extreme modeling assumptions; discrepancies between the simulated ground truth and the used models are systematically varied and they are greater than the discrepancies between the used models and the actual processes.

The system is evaluated with a special testing harness, where it is fed with synthetic data obtained with the help of simulated plumes and detector activations. The simulation assumes a specific situation, such as for example a uniform grid of detectors and a set of 5 potential sources positioned along a line in the middle of the grid shown in Fig. 7. The simulated plume originates from the source at location 3 and its shape is assumed to be triangular cone. The area enclosed by the plume shape represents the region where the concentrations exceed the detection sensitivity and it is also used for the data association; all the observations collected inside the cone are considered for hypothesis updating.

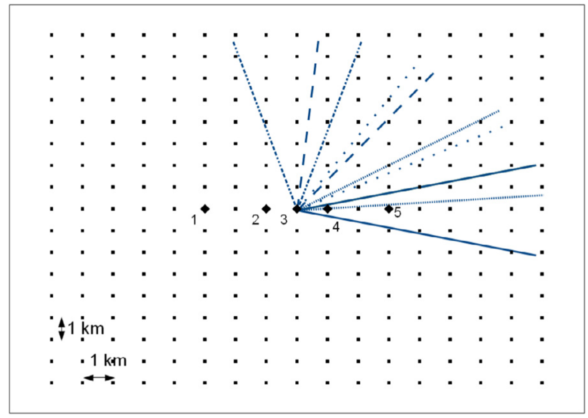


Fig. 7. Experimental set up. Dots represent sensors and diamonds represent potential sources. Lines with the same pattern represent the plume boundaries.

The simulated environment is used to sample values of the detectors under different conditions. The sampled data is fed to the distributed inference system and the expected localization accuracy is estimated (see [4] for a detailed discussion). The average localization accuracy of the overall system is plotted in performance charts as a function of the detector noise. Fig. 8 shows a performance chart where each curve represents the expected localization accuracy for grid densities 1 km, 0.5 km and 0.25 km, respectively. The value on the vertical axis represents the percentage of the cases in which the correct hypothesis was associated with the maximum posterior and the plume direction was assumed 180°, i.e. the plume propagated towards the North, thus perpendicular to the sources line.

The performance chart in Fig. 8 shows good accuracy also in the case of high noise levels. Moreover, the charts show that the expected accuracy improves with the detector grid density, a theoretically predicted property of the used causal models [14]. Namely, by increasing the detector grid density the number of branches in the underlying domain model increases which mitigates the impact of the sensory noise and the modeling discrepancies. Similar plots can be made for different orientations [4]. Moreover, the impact of various simplifying assumptions was evaluated through similar experiments. For example the impact of the errors made with the assumptions about the plume form used for the data association was evaluated by running experiments with different shapes of assumed and true plumes. These experiments showed that the localization system is insensitive to the geometry of the plume (see [4]).

The introduced performance charts in combination with robustness evaluations are engineering tools that support systematic design of cost efficient solutions with known expected accuracy. We can use such tools to investigate trade-offs between the sensor network density and the sensor quality that guarantees a lower bound on the expected performance of the overall system.

6. A visual interface for augmented human smell perception

Humans are very good at detecting and discerning smells [22]. However, we are bad at identifying the name of a smell or at providing a meaningful description of it. Olfactory perception is

[3] Note, in this paper gas detectors can be devices or human observers. In case of human sources, it is assumed that they are independent in the early stage of an incident, as the people might be biased through communication via social media in later phases.

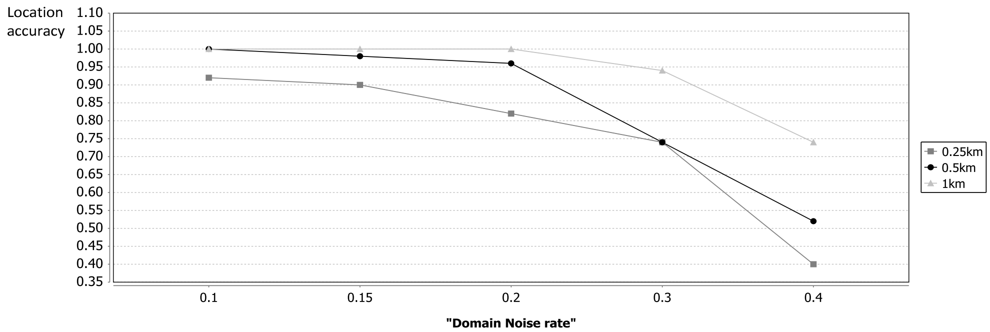


Fig. 8. Expected localization accuracy as a function of the detector noise rate (i.e. failure rate). The different curves correspond to different sensor densities.

an interesting, yet underrepresented research area in human–computer interaction (HCI) and there has been little research on how to create interfaces making use of or supporting the human sense of smell.

To incorporate human smell capacity into an environmental monitoring system, a dedicated mobile application was developed. In order to create such a dedicated mobile application, we investigated the effect of multi-modal odor cues on human smell identification. Our focus was on providing a mobile interface that assists users in identifying the smells they perceive to increase the accuracy and performance of the overall environmental monitoring system.

Olfactory perception in HCI research

One of the main obstacles in research on olfactory interfaces (discussed in [23]) concerns the lack of proper classification or description schemes for smells (other than those related to wine and perfume). The lack of a common smell 'vocabulary' makes communicating about smells difficult for both system and user.

Another difficulty is that although humans can detect and discriminate countless odorants, they can identify few by name [22]. For a public environmental monitoring system based on human/social sensing this is problematic because the system is crucially dependent on accurate smell descriptions from humans. Therefore, we studied whether there are suitable ways to facilitate the communication of sensory information, such as olfactory perceptions.

Previous research has shown that the identification of odors (i.e. free recall of the name of an odor) without the presence of other relevant semantic information is a difficult task for humans (e.g. [24], [25]). Other studies have revealed that the odor identification process works considerably better when participants can choose between possible labels, rather than free recall (e.g. [26]).

Yeshurun and Sobel [22] argue that the primary function of olfaction can be viewed as to signal a human for approach or withdrawal and that this signal is best represented by pleasantness. Their research indicates that humans consistently and rapidly describe an odor by its perceived pleasantness. Gottfried and Dolan [27] demonstrate that although human olfaction is unreliable, it benefits substantially from visual cues. Participants were faster and more accurate at recognizing odors when these odors appeared in the context of semantically congruent visual cues. Morrot et al. [28], Demattè et al. [29], and Gilbert et al. [30] all found strong connections between olfactory perception and colors.

To summarize, the literature suggests that odors are subjectively associated with images, words, colors and personal memories. Furthermore, people tend to discriminate between smells based on how pleasant or unpleasant they find the smell.

Stimuli to assist odor recognition

To create a first prototype of an intelligent interface that automatically generates cues to assist human smell identification for this application, our question is:

"*Which on-screen stimuli will improve the accuracy of user provided, real-time odor identifications compared with free recall?"*.

We carried out a study in two phases. In phase one, we elicited odor descriptions and associations in order to extract meaningful stimuli for human smell identification as well as to see which stimuli provide discernible rankings for automatic smell classification. In phase two, we carried out an experiment to assess which smell-related stimuli generated by the interface would lead to more accurate odor identification. Details of these studies can be found in [31].

In phase 1 of the study, we exposed participants to nine distinct smells and asked them to provide multiple types of associations that the smell evoked. The between-subjects independent variable 'odor' was manipulated by exposing each participant to one of nine odors with 'scratch & sniff' cards specifically developed by a professional manufacturer for the purpose of the study. The final sample consisted of 429 participants (180 males and 249 females). Ages ranged from 18–65 ($M = 33.1$, $SD = 12.46$).

Each participant was asked to provide

1. **a textual association**: 'Which word or term best describes the current smell?';
2. **a personal memory association**: 'Describe a memory you associate with the current smell';
3. **a color association**: a choice from black, white, red, green, yellow, blue, brown, purple, pink, orange, gray colors provided on the screen;
4. **a visual mental image the odor evoked**: 'Describe a visual image that comes to mind when you smell the odor'.

For each association method, we measured how confident participants' felt that their provided association accurately described the odor. Each time participants provided an association, they were therefore asked to answer the following questions:

Confidence/closeness of match: Participants rated on a seven point Likert-type scale how confident they were that the selected association best matched the odor they perceived (e.g., 'How confident are you that the image accurately describes the odor?'; scale from 1–7, 1 = 'not a close description at all' to 7 = 'very close description').

Participants then also reported their perception of intensity and pleasantness of the smell.

Intensity: Participants rated on a seven point Likert-type scale the perceived strength of the odor ('How intense was the odor?').

Pleasantness: Participants rated on a seven point Likert-type scale the perceived pleasantness of the odor ('How pleasant did you find the odor?').

In a short post-test interview, participants were also asked which association method they thought best described the smell and why.

The results showed that there were indeed significant differences between how intense a smell was perceived depending on which odorant participants were rating ($H(8) = 49.45, p < 0.001$). 'Natural gas' was perceived as the most intense smell ($M = 5.7, SD = 1.3$) and 'Mildew' as the least intense one ($M = 4.2, SD = 4.2$).

The results of another analysis showed that there were also significant differences between pleasantness ratings ($H(8) = 101.89, p < 0.0001$) depending on which odorant participants were rating.

We then analyzed whether there were (significantly) different odor associations for each association method. If odor associations differed significantly across odors for a particular method (e.g., image associations) then that association method would be best to implement as stimuli in our system. We found that there was indeed a significant difference between color associations ($H(8) = 19.69, p = 0.012$) depending on the odorant participants were confronted with. Thus, color associations were a good candidate to be used for providing feedback regarding olfactory perceptions. However, looking at the frequencies of provided color associations, we found that although some of the provided color associations fitted well with the odorant, 'brown' was the most often selected color for six of the nine smells, which made color cues much less meaningful. Moreover, even though the image and name descriptions were very different for the different smells, there was too much variation in the answers to yield significance. We did not find significant differences between odors in terms of memory valence.

Further results show that there was a significant interaction effect between participants' confidence ratings on the association methods and the different odors ($F(31.26, 0.98) = 1.70, p = .009$), meaning that confidence levels differed significantly for each method for a particular odor. Since we wanted to offer stimuli that will yield highly accurate smell descriptions, it was important to us that users would feel confident about the description/association they provide.

To assess the quality of a particular association method, we compared participants' confidence rating (how well the participant felt the association matched the odor) and found that overall participants felt least confident with the memory association (29.9%), followed by the image (26.9%), text (25.1%), and, finally, their color association (18%). Similarly, from the interview responses we found that participants were most confident about their memory associations (31.9%), followed by image associations (27.9%), textual associations (25.1%), and least confident about color associations (18.0%).

To summarize, we found that participants rated the odors differently in terms of intensities and pleasantness, again demonstrating that human noses are able to distinguish between and characterize different odors. Both survey and interview data showed that participants were most confident of their memory, image and text associations. Memory associations were individually very different and particularly difficult to present. Therefore, we decided to focus on image, text, and a combination of image and text cues, as well as pleasantness ratings in a second study to test whether and which of these stimuli types would lead to the highest accuracy and user satisfaction.

In phase 2 of the study, we carried out an experiment to assess whether the association stimuli elicited from phase 1, automatically generated by an adaptive interface, would lead to more accurate odor identification. The final sample consisted of 190 participants (106 males and 84 females). Ages ranged from 18–65 ($M = 32.0, SD = 10.60$).

We focused on the three smells we considered to be most common and realistic in the environmental monitoring context: gasoline, rotten eggs, and natural gas. Also, there was considerable overlap in the descriptions received from participants in phase one for these smells and we therefore expected these smells to benefit most from additional stimuli to aid recognition (see Table 2).

Table 2
Study 2 odor name stimuli.

Odor	Association	'Similars'
Natural gas	Stove	e.g., rotten eggs, exhaust fumes
Gasoline	Petrol station	e.g., paint, oil
Rotten eggs	Rotten eggs	e.g., rotten vegetables, sweat

Each participant was exposed to only one of five on-screen stimulus conditions:

- Textual: The participant could select a name that best matched the odor from a list of 25 words.
- Image: The participant could select an image that best matched the odor from a list of 25 images (see Fig. 9).
- Text/image: The participant could select a text/image that best matched the odor from a list of 25 text/image associations.
- Pleasantness: The participant could provide feedback on a 7-point pleasantness scale: 'How pleasant do you think this odor smells?'.
- Free recall: Participants were asked to fill in the name of the odor in a textbox. They were not provided with any additional stimuli.

The between-subjects design meant that a participant would be exposed to only one stimulus (for instance images), he or she first smelled the odor, and was then exposed to all the available image cues so that they could click on the figure that best represented the smell they perceived.

In the environmental monitoring system that is developed, the system would generate hypotheses about potential gases in the environment, each associated with a calculated probability. The users' odor identifications would help to reduce the number of hypotheses and determine the gas that is in the air. This study is designed to closely replicate such a situation.

A new variable was calculated to measure user accuracy: two researchers coded all user responses. An 'accurate' score was assigned to every selection that directly represented the odorant (e.g., 'Rotten eggs' for rotten eggs) or through selection of 'similar descriptors' as identified from phase 1 (e.g., 'Oil' for 'Gasoline'). An 'inaccurate' score was assigned when the odor was not identified.

Results showed a significant difference in accuracy between free recall and the stimulus conditions combined ($U = 1906.00, Z = -1.982, p = 0.048$). When comparing accuracy means ($1 =$ direct; $2 =$ 'similar', and $3 =$ wrong), participants were most accurate in the image stimuli condition ($M = 1.35, SD = .49$), the image and word combined condition ($M = 1.42, SD = .50$), and then the word only condition ($M = 1.54, SD = .50$). They were least accurate in the free recall condition without any stimuli ($M = 1.63, SD = .49$).

The results of our studies indicate that interactive stimuli indeed improve human smell identification performance. The results show that participants thought that providing an image and word stimulus was most pleasant and useful. Moreover, we found a significant difference in accuracy between free recall and the stimulus conditions combined. No significant difference in accuracy was found for the 4 stimulus conditions. When comparing accuracy, participants were most accurate in the image stimuli condition, the image and word combined condition and then the word only condition. They were least accurate in the free recall condition without any stimuli.

Development of a visual interface supporting human olfactory perception

Based on the results of the previous study we were able to develop an adaptive interface of a mobile application for the Android operating system. This application is connected to a

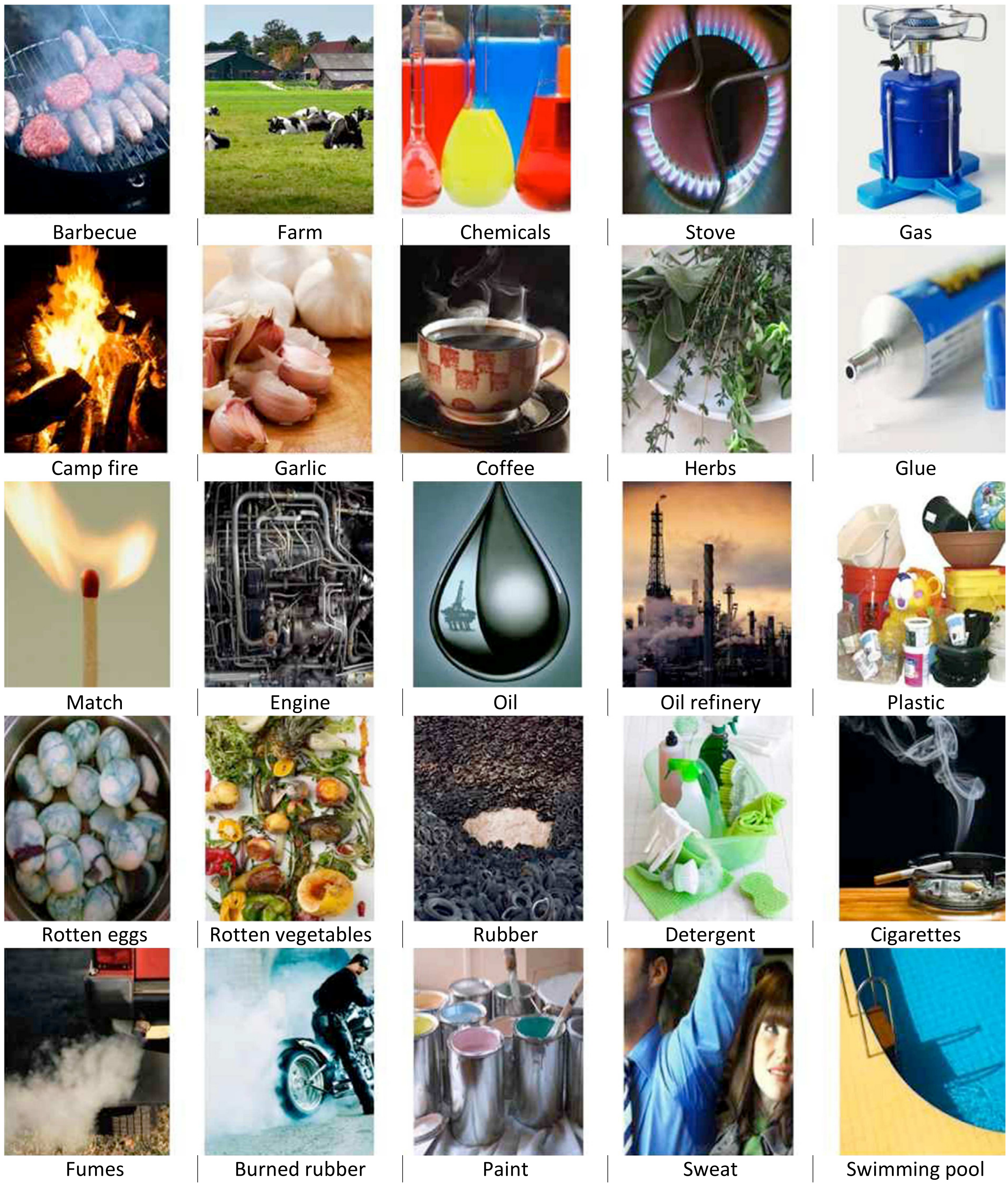


Fig. 9. Visual cues presented in 2nd study.

distributed environmental monitoring and decision-making system. The environmental monitoring system detects anomalies in the air quality through chemical sensors or complaints from inhabitants. It consequently calculates hypotheses concerning the gases that are the most likely pollutants. Inhabitants in the area are contacted through their mobile phones and requested about potentially perceived smells in order to inform the detection system and to eventually support or reject hypothesis based on this newly received evidence. The adaptive interface on participating volunteers' mobile phones dynamically generates and displays visual and textual cues related to the hypothesized gases. The user can then select the image/text combination that most accurately represents the smell they perceive.

For our application we developed a Bayesian Network, which contains the relations between the presence of the three gases H_2S, mercaptan and oil/gasoline and the observation of answers to three types of questions illustrated in Fig. 10. The questions we use for classification make use of a word/image combination, ratings of the smell on a pleasantness scale and classifying the gas into a category.

Within the context of the overarching environmental monitoring project the goals of the mobile application are twofold: First,

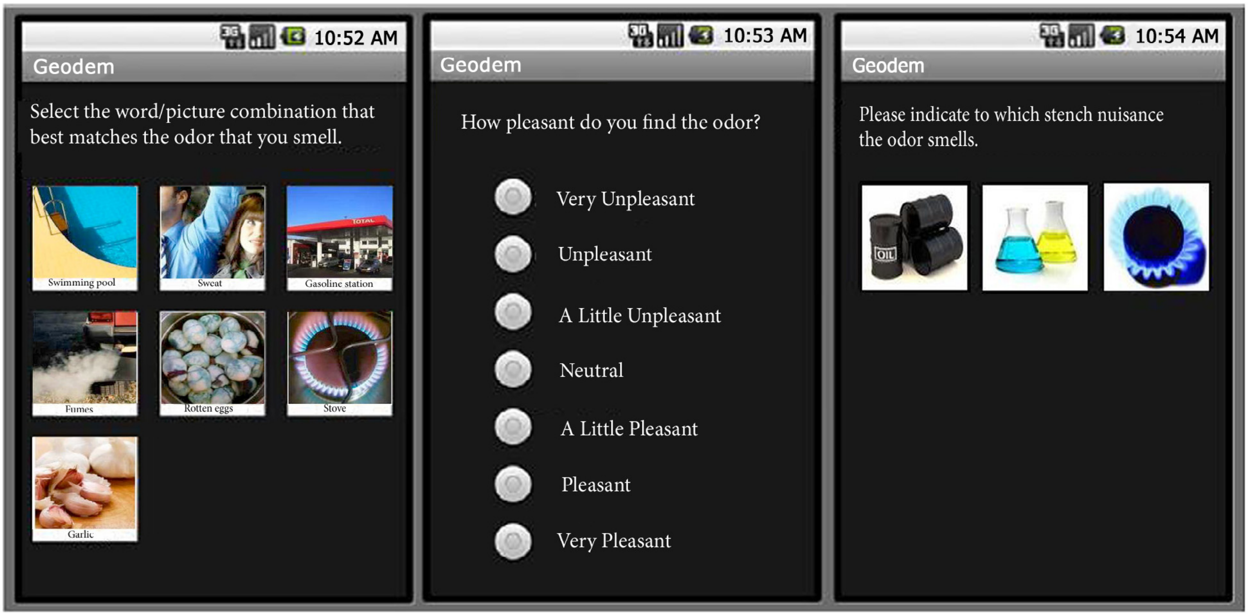


Fig. 10. Screenshots of the prototype app.

concerned citizens can file a complaint and inform responsible environmental monitoring agencies about unusual and unfavorable smells at their current location. Second, the system can inquire about users' smell perceptions in a potentially affected area, which are then communicated to the central detection system, in order to determine the likelihood, severity, and location of an incident.

The interface offers users one of three stimulus screens (see Fig. 10) based on which kind of feedback promises the highest information gain for the underlying Bayesian Network-based gas detection system. (1) A screen which shows visual and textual cues, as tested in this study. (2) A 7-point pleasantness scale to indicate the subjective pleasantness of the perceived odor and (3) a screen allowing to select whether the user perceives the smell as oily, chemical, or gaseous (which allows the gas detection system to make critical computations concerning the likelihood and criticality of a gas pollution). The selection of stimuli is based on Bayesian reasoning processes concerning possible substances derived from already available information from stationary sensors or earlier reports from concerned citizens.

As a proof of context implementation, the application is based on a simple HTTP client–server architecture implemented with the Web framework Django (www.djangoproject.com) and a trial version of the Bayesian decision-network tool HUGIN (www.hugin.com) on the server side, and Android (www.android.com) on the client side.

The conditional probabilities entered in the Bayesian Network are mostly approximations, whereas some stem from unpublished research. By iterating over all the leaf nodes of our Bayesian Network, which represent (possible) answers given by our users, we can determine which question is going to give us the most discriminative information. This is done by 'clamping' a node, which means selecting a possible state/answer for that node, and observing the difference in probability for the states in the topmost node. The leaf node that shows the biggest difference and thus is the most informative is then sent back to the users in the form of a question. By answering that question, the user effectively permanently 'clamps' this node and all the probabilities in the network are updated. When asked for the most informative node again, the system will return a new (unclamped) node which, given the state the network is in, is now the most informative. This process starts up every time a person indicates that there has been an incident and via this question selection scheme speeds up the gathering of useful information. This is even more the case when many different questions are added. Also the concept of using tailor-made questions for users seems powerful. For instance, different word/picture combinations can be presented to a user when the system requires information about some particular examples.

Based on these findings, we built an initial prototype of a mobile interface that allows users to report unusual and/or unpleasant odors via a combination of visual and textual cues. We expect that this kind of 'social environmental sensing' will considerably improve early detection of environmental incidents in the near future.

To our knowledge, the study reported here was the first large-scale user study to inform the development of a system to facilitate human smell perception and labeling in a mobile application. We have shown that computer-supported odor classification benefits from particular types of graphical user interface stimuli. Both perceived enjoyment and usefulness were highest for the picture plus word stimuli condition, which is a crucial finding for mobile applications designed for environmental monitoring that heavily relies on participation from volunteers. Interestingly, we also found that pleasantness is indeed a differentiating indicator of smell perception. In our studies pleasantness ratings differed significantly across smells.

7. Twitter as real-time stench locator

The use of social media gives new additional ways to detect and locate stench. Twitter has been shown to be a suitable source of certain information as there are over 300 million twitter accounts and the communication is almost real time. In addition, many tweets are tagged with geographical information. There are already several approaches proposed to use twitter as a monitoring tool, such as in event detection of flu symptoms in a population [32] or for earth quake alerts [33].

Since many stench related keywords are part of daily life simple twitter mining is not sufficient. We developed a method [34] that calculates a probability score for the detection of anomalies and that also takes into account social factors, such as day and night rhythm and geographical distribution of users.

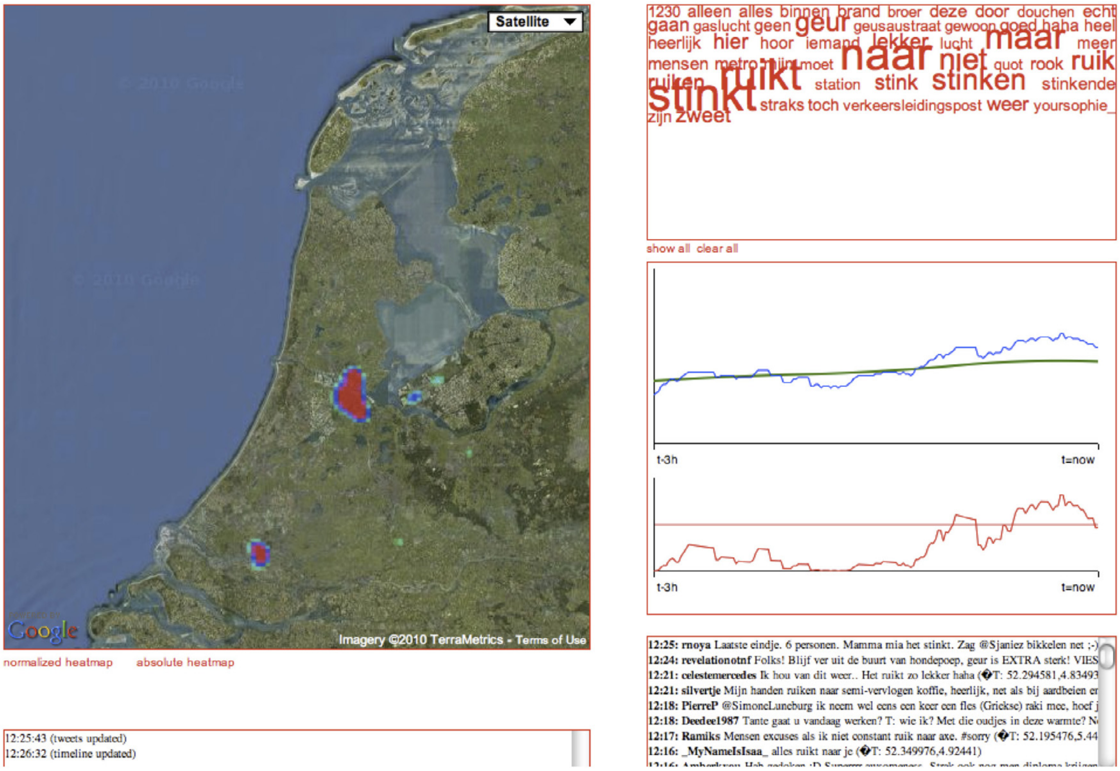


Fig. 11. This figure shows on the left a screenshot of the interface. On the right are two graphs of stench related tweets. The upper graph shows the actual number of tweets as well as the expected number. In the lower graph the difference between the two is visualized. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

In this work, we monitor all tweets that contain stench related keywords in a certain area. To detect anomalies we have to model the expected spatio-temporal twitter activity of stench related keywords when no incident happens. The differences between the actual and the expected numbers of stench related tweets forms the basis for the detection of anomalies and the highlighting of those differences for real-time visualization of potential events.

We model the occurrences as a non-stationary Poisson process, i.e. a Poisson process with a varying rate parameter. In our case we model the tweet rate λ as a function of both space and time. For a certain time interval the expected number of observations can be obtained by integrating over the time interval and area.

We are interested in the probability of the number of observations our model generates in a 'normal' situation plus an unknown number of incident related tweets. We gauge this by calculating the probability that the observations are caused by a model with a higher rate parameter λ. For any time window and area this is calculated based on the following formula:

$$P(anomaly|n, \lambda_M) = \int_{\lambda_M}^{\infty} Pois(\lambda, n)\mathrm{d}\lambda = 1 - \int_{0}^{\lambda_M} Pois(\lambda, n)\mathrm{d}\lambda. \quad (3)$$

Here, λ_M is the expected rate derived from our space–time model M, n is the observed number of tweets, and $Pois$ is the Poisson distribution's probability mass function.

A set of 93 predefined queries of pollution related keywords and phrases were issued to a Twitter search engine. The keywords were obtained from complaint logs of a regional environmental management agency from a list of human smell descriptions of industrial gases. If exact geographical coordinates were available (which is the case for many tweets sent from current smart phones), those were used. Otherwise, the filtering was based on user-provided location labels, which we transformed into geographical coordinates using Google's Geocoding API. Over 120.000 tweets containing one or more 'smelly' keywords from over 50.000 users in NL were obtained during the data collection period. Of these tweets, 72% had a location from which 44% was exact and 55% approximate.

Locations of relevant tweets are plotted on a map that is centered and focused on the Netherlands: the location of interest. Visualizing spatial data on a two dimensional plane allows for detection of correlations and clusters in the patterns formed by the data points. High densities of relevant tweets at a certain location provide users of the system with a first indication of potential anomalies. The map is augmented with a visual representation of the anomaly: for every location, the probability of an anomaly is translated into a color value, ranging from green (very improbable) to red (highly probable), as can be seen in Fig. 11. The resulting image is added as a semi-transparent overlay to the map. The upper figure on the right indicates the actual number of relevant tweets in blue and the expected number of tweets in green. In the lower figure the difference is shown between the actual and the expected number of tweets together with a threshold above which an anomaly is detected.

As no major incident happened during the data collection period, we manually collected a random sample of tweets and compared that to actual pollution incidents The preliminary evaluation revealed that while most tweets containing pollution-related keywords did not actually report 'serious' visual or olfactory observations, two major fires were indeed detected and highlighted on the system's map. This illustrates that the system is capable of detecting pollution-related anomalies. However, there were also many occasions where the system did not detect incidents. Although this preliminary system can be improved in number of aspects it shows the potential of using social media data as an additional source of information in environmental systems.

8. Conclusions

Real world problems such as environmental monitoring require a close cooperation between humans and intelligent agents. In these systems a human does not only interact with the environment and is a user of information but also forms an essential source of information. The challenge is to design robust distributed systems that fuse heterogeneous and dynamic information from humans and intelligent agents to support decision making. In this paper we focus on two essential aspects of such systems: the robustness of information fusion and decision making and optimal human interaction. The application of this paper is a system to monitor air quality and to detect environmental incidents in an urban area.

Modular Bayesian Networks in conjunction with Inference Meta Models were presented as a solid basis for the development of efficient and reliable fusion systems, which support situation assessment in environmental monitoring. It creates the possibility to add and delete modules on the fly resulting in a kind of data-driven fusion. This gives the possibility to deal with ad-hoc constellations of information sources as it is difficult to provide adequate domain models prior to the operation in these types of applications If many information sources are available, we can obtain BNs corresponding to factor trees with large branching factors which make fusion reliable even if we use parameters that deviate from the true distributions significantly.

As this paper concerns the detection and tracking of gases, the human interface is an interface for human olfactory perception. A mobile application was developed that assists users in identifying the smells they perceive to increase the accuracy of the overall environmental monitoring system. In order to create such a dedicated mobile application we investigated the effect of multi-modal odor cues on human smell identification. Our experiments show that both a combination of text and picture descriptors and individual pleasantness ratings are useful to accurately identify smells when the number of smells is limited (as it is the case in an industrial area) and the sample size is large enough.

Promising is also the use of social media. We showed in a preliminary study that stench related tweets form an additional source of information to detect environmental incidents.

In this paper we presented a concept of which the components have been validated but still an effort of integration and extensive testing is needed to see how the distributed agent–human system organizes itself and works together. In particular with respect to the primary user of the system: the environmental agency, which forms the 'spider in the web'. This agency is responsible for the handling of chemical incidents and decides about the measures to be taken. They are also responsible for the deployment of sensors and can involve experts in the field in case of an chemical incident as additional sources of information.

The application of this paper is in the domain of environmental protection and a human–agent system for monitoring air pollution. However, the solutions can be generalized to other domains were (sensor) agents and humans share and fuse information in a distributed setting. Potential other application domains are for instance dynamic traffic routing, fire monitoring and dynamic selecting escape routes, weather monitoring and local road conditions, search and rescue.

One contribution of this work is the presented Modular Bayesian Networks and Inference Meta Models, which offer reliable dynamic fusion when modules come and go. So it is easy to fuse human information where humans supply information on the fly. Another contribution is the developed human interaction with the system. The smell study reported here was the first large-scale user study to inform the development of a system to facilitate human smell perception and labeling in a mobile application.

The work reported in the paper has laid the basis for robust information fusion in environmental management when humans and sensor information has to be dynamically fused. Future research should involve full integration and large scale testing of the system. Also dynamic selection of places and persons from which the information will decrease the uncertainty most is an important issue.

The impact on environmental management of these type of intelligent systems can be substantial, as currently it is common practice in agencies to handle incoming complaints through phone calls by hand.

Acknowledgments

The presented work was partially funded by the European Union, FP7-ICT project DIADEM, ref. no. 224318 and the Dutch Ministry of Economic Affairs, ICIS project ref. no. BSIK03024.

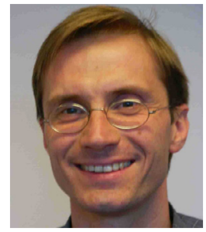
References

[1] A. Winterboer, M.A. Martens, G. Pavlin, F.C.A. Groen, V. Evers, DIADEM: a system for collaborative environmental monitoring, in: Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work, 2011, pp. 589–590.
[2] S. Asadi, M. Reggente, C. Stachniss, C. Plagemann, A.J. Lilienthal, Statistical gas distribution modelling using kernel methods, in: E.L. Hines, M.S. Leeson (Eds.), Intell. Syst. for Machine Olfaction: Tools and Methodologies, IGI Global Pub., 2011, pp. 153–179 (Chapter 6).
[3] A.J. Lilienthal, S. Asadi, M. Reggente, Estimating predictive variance for statistical gas distribution modelling, in: Proc. Int. Symp. Olfaction and Electronic Nose-2009, in: AIP Conf. Proc., vol. 1137, 2009, pp. 65–68.
[4] F. Mignet, G. Pavlin, P. de Oude, P.C.G. da Costa, Evaluating complex fusion systems based on causal probabilistic models, in: FUSION 2013, 2013, pp. 1590–1599.
[5] G. Pavlin, P. de Oude, F. Mignet, Gas detection and source localization: A Bayesian approach, FUSION 2011, 2011, pp. 1–8.
[6] J. Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann, Burlington, 1988.
[7] A. Keats, E. Yee, F.-S. Lien, Bayesian inference for source determination with applications to a complex urban environment, Atmos. Environ. 41 (3) (2007) 465–479.
[8] U. Konda, Y. Cheng, T. Singh, P. Scott, Source identification of puff-based dispersion models using convex optimization, in: Proceedings of the 13th Conference on Information Fusion, UK, July, 2010.
[9] K.S. Rao, Source estimation methods for atmospheric dispersion, Atmos. Environ. 41 (33) (2007) 6964–6973.
[10] G. Pavlin, P. de Oude, M. Maris, J. Nunnink, T. Hood, A multi-agent systems approach to distributed Bayesian information fusion, Inf. Fusion 11 (3) (2010) 267–282.
[11] F.R. Kschischang, B.J. Frey, H.-A. Loeliger, Factor graphs and the sum–product algorithm, IEEE Trans. Inf. Theory 47 (2001).
[12] M.A. Paskin, C.E. Guestrin, Robust probabilistic inference in distributed systems, in: Proceedings of the 20th Annual Conference on Uncertainty in Artificial Intelligence, UAI-04, Banff, Canada, AUAI Press, 2004, pp. 436–445.
[13] Y. Xiang, Probabilistic Reasoning in Multiagent Systems: A Graphical Models Approach, Cambridge University Press, Cambridge, 2002.
[14] G. Pavlin, J. Nunnink, Inference meta models: Towards robust information fusion with Bayesian networks, in: The Ninth International Conference on Information Fusion, Florence, Italy, 2006.
[15] P. de Oude, G. Pavlin, Efficient distributed Bayesian reasoning via targeted instantiation of variables, in: IAT 2009, 2009, pp. 323–330.
[16] N. Jennings, K. Sycara, M. Wooldridge, A roadmap of agent research and development, Auton. Agents Multi-Agent Syst. 1 (1998) 7–38.
[17] R.A. van Engelen, Approximating Bayesian belief networks by arc removal, IEEE Trans. Pattern Anal. Mach. Intell. 19 (8) (1997) 916–920.
[18] S.P. van Gosliga, F.C.A. Groen, Estimating the impact of variables in Bayesian belief networks, in: Tenth Tbilisi Symposium on Language, Logic and Computation: Gudauri, Georgia, 23–27 September 2013, Tbilisi: Centre for Language, Logic and Speech, and Razmadze Mathematical Institute, Tbilisi State University, 2013, pp. 87–90.
[19] E. van Foeken, P. Hiemstra, L. Kester, An improved method for creating shared belief in communication constrained sensor networks, in: 4th German Workshop Sensor Data Fusion: Trends, Solutions, Applications SDF2009, Lübeck, Germany, 2009.
[20] R.O. Duda, P.E. Hart, Pattern Classification and Scene Analysis, John Wiley & Sons, New York, 1973.
[21] C. Bishop, Pattern Recognition and Machine Learning, Springer-Verlag, Berlin, 2006.

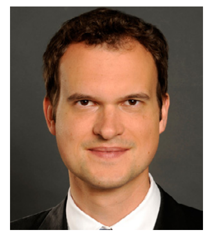
[22] Y. Yeshurun, N. Sobel, An odor is not worth a thousand words: From multidimensional odors to unidimensional odor objects, Annu. Rev. Psychol. 61 (2010) 219–241.
[23] J. Kaye, Aromatic output for HCI, Interactions 11 (1) (2004) 49–61.
[24] W.S. Cain, To know with the nose: Keys to odor identification, Science 203 (1979) 467–470.
[25] C. Murphy, W.S. Cain, M.M. Gilmore, R. Blair Skinner, Sensory and semantic factors in recognition memory for odors and graphic stimuli: Elderly versus young persons, J. Psychol. 104 (2) (1991) 161–192.
[26] R. de Wijk, W.S. Cain, Odor quality: description versus free and cued identification, Percept. Psychophys. 56 (1) (1994) 12–18.
[27] J. Gottfried, R. Dolan, The nose smells what the eye sees: Cross-modal visual facilitation of human olfactory perception, Neuron 39 (2) (2003) 375–386.
[28] G. Morrot, F. Brochet, D. Dubourdieu, The color of odors, Brain Lang. 79 (2) (2001) 309–320.
[29] M.L. Demattè, D. Sanabria, C. Spence, Cross-modal associations between odors and colors, Chem. Senses 31 (2006) 531–538.
[30] A.N. Gilbert, R. Martin, S.E. Kemp, Cross-modal correspondence between vision and olfaction: The color of smells, Am. J. Psychol. 109 (3) (1996) 335–351.
[31] A.K. Winterboer, V. Evers, F.C.A. Groen, G. Pavlin, A Visual Interface for Augmented Human Olfactory Perception in the Context of Monitoring Air Quality, Technical report, University of Amsterdam, 2011.
[32] V. Lampos, N. Cristianini, Tracking the flu pandemic by monitoring the social web, in: Workshop on Cogn, Inform. Proc., CIP'10, 2010, pp. 411–416.
[33] P. Earl, Earthquake twitter, Nature Geosci. 3 (2010) 221–222.
[34] H. Smid, P. Mast, M. Tromp, A. Winterboer, V. Evers, "Canary in a coal mine": monitoring air quality and detecting environmental incidents by harvesting Twitter, in: Proceedings of the 2011 Annual Conference Extended Abstracts on Human Factors in Computing Systems, 2011, pp. 1855–1860.

Frans Groen is professor emeritus of the University of Amsterdam. He obtained his M.Sc. degree (cum laude) and his Ph.D. degree in Applied Physics from the Delft University of Technology. From 1970 to 1988 his was a scientific staff member of the Pattern Recognition Group at the Applied Physics Department of the TU-Delft, where he was responsible for the robotics activities. In 1988 he became full professor at the Computer Science Department of the University of Amsterdam, heading the research in Intelligent Autonomous Systems. He served as cross-appointed full professor at the Free University of Amsterdam from 1986 to 1996. He was a Fulbright research scientist in 1984 at the Robotics Institute at CMU and in 1996 as visiting professor at the University of Utah. He was an advisor of TNO D&V, the National Dutch Research Lab for Defense and Safety. From 2001 to 2007 he served as director of the Informatics Institute of the University of Amsterdam. He is a fellow of the IEEE.

His current research interests are in sensor data processing for autonomous real-world multi-agents systems, dynamic distributed world modeling and decision making.



Gregor Pavlin is senior researcher and project manager at Thales Nederland B.V. He received his Ph.D. in computer science from the Graz University of Technology, Austria in 2001. He has extensive industrial experience in development of complex, safety critical airborne software systems. His current research interests are algorithms, modeling techniques and architectures supporting robust decentralized information fusion with humans in the loop. The emphasis is on the exploitation of causal Bayesian models for the development of sound solutions using heterogeneous data from noisy sources.



Andi Winterboer is a senior consultant at VDI/VDE Innovation + Technik GmbH in Berlin. VDI/VDE Innovation + Technik GmbH, is a service provider for the German Federal Ministry of Education and Research, where he is responsible for technology screening on all topics related to human–computer interaction. He obtained his Ph.D. in Artificial Intelligence from the University of Edinburgh in 2009. Before joining VDI/VDE Innovation + Technik GmbH in 2011 he was a postdoctoral researcher at the Intelligent Systems Lab (ISLA) at the University of Amsterdam participating in the EU FP7 project DIADEM. Within this project, he worked with Professor Vanessa Evers on human–computer interaction aspects of a system that detected potentially hazardous situations in heavily populated industrial areas using input from both a distributed sensor network and humans through mobile devices.

Vanessa Evers is a full professor of Computer Science at the University of Twente Human Media Interaction group and Director of the DesignLab for multidisciplinary research. She received a M.Sc. in Information Systems from the University of Amsterdam, and a Ph.D. from the Open University, UK. During her Master studies she spent two years at the Institute of Management Information Studies of the University of New South Wales, Sydney. After her Ph.D. she has worked for the Boston Consulting Group, London and later became an assistant professor at the University of Amsterdam Institute of Informatics. She was a visiting researcher at Stanford University (2005–2007). Her research interests focus on interaction with intelligent and autonomous systems such as robots or machine learning systems as well as cultural aspects of Human–Computer Interaction. She has published over 80 peer reviewed publications, many of which in high quality journals and conferences in human–computer interaction and human robot interaction.