International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:4, No:1, 2010

Context Generation with Image Based Sensors: An Interdisciplinary Enquiry on Technical and Social Issues and their Implications for System Design

Julia Moehrmann, Gunter Heidemann, Oliver Siemoneit, Christoph Hubig, Uwe-Philipp Kaeppeler, Paul Levi

***Abstract*—**Image data holds a large amount of different context information. However, as of today, these resources remain largely untouched. It is thus the aim of this paper to present a basic technical framework which allows for a quick and easy exploitation of context information from image data especially by non-expert users. Furthermore, the proposed framework is discussed in detail concerning important social and ethical issues which demand special requirements in system design. Finally, a first sensor prototype is presented which meets the identified requirements. Additionally, necessary implications for the software and hardware design of the system are discussed, rendering a sensor system which could be regarded as a good, acceptable and justifiable technical and thereby enabling the extraction of context information from image data.

***Keywords*—**Context-aware computing, ethical and social issues, image recognition, requirements in system design.

I. Introduction

CONTEXT-AWARE systems are of growing interest in the last years. According to classical definition by Anind K. Dey et al., context is "any information that can be used to characterize the situation of an entity" [1]. Context-aware applications and/or systems are therefore information and communication systems which adapt their behavior according to the situation they are in. Since acquiring and modeling context information is a tedious and expensive task, it is beneficial to share context information between different applications or systems [2, 3, 4, 5]. So-called spatial models are intended to represent or mirror certain aspects of the real world as closely as possible, thereby serving as a shared, common basis for different context-aware applications. Spatial models, however, at the same time raise the challenge of how to update them if the environment changes [6, 7]. Besides explicit modeling by the system designer, most of the changes are amenable by different kinds of sensors such as temperature, lighting, pressure or acceleration sensors. In this paper, however, we want to concentrate on a special kind of "sensors" – IP cameras – and how context can be generated from image data. The application of cameras as sensors could replace a series of other physical sensors and the need to augment the physical world would decrease and thereby simplify the creation of context-aware applications dramatically. This idea becomes increasingly clear if one bears in mind that there are environments, e.g. in manufacturing, where the proliferation of sensors or sensor nets, due to the extreme physical conditions such as heat or electromagnetic inference fields, is not possible at all [8]. Also the proliferation of so-called Smart Dust is considered in certain realms as a kind of environmental pollution [9]. There, *image based sensors*, i.e. sensors that generate context from image data, seem to remain the only real alternative for acquiring important context information. Moreover image-based sensors seem to be very attractive since they are also very flexible and easily extensible. Simply by changing the image or pattern recognition method, the given sensor could be turned into another kind of sensor which gathers different context information for different purposes. Image based sensors are therefore considered to be very versatile, easily reusable and thus quickly deployable in different fields of application.

As already indicated, the technical basis for acquiring context information from cameras, i.e. image based sensors, is the analysis of the acquired image data by certain pattern recognition algorithms. This analysis of image data can, today, be easily performed with the aid of a variety of existing image recognition systems. However, currently available image recognition systems have some major drawbacks: First, current systems are highly dependent on the application, i.e. on the specific recognition task. Second, the creation of such recognition systems demands expert knowledge. An unskilled user without prior knowledge about pattern recognition methods is neither able to define an object recognition task on a technical level nor is willing to spent a lot of time on training the recognition system.

To solve this problem, we developed a framework for the user-defined creation of image recognition systems, which allows for the easy and quick extraction of context information from image data. However, setting up cameras either in the public, at the workplace or in private areas, at the

Julia Moehrmann and Gunter Heidemann are with the Institute for Visualization and Interactive Systems, Intelligent Systems Department, University of Stuttgart, Germany, {firstname.lastname}@vis.uni-stuttgart.de

Oliver Siemoneit and Christoph Hubig are with the Institute of Philosophy, Chair for the Philosophy of Science and Technology, University of Stuttgart, Germany, {firstname.lastname}@philo.uni-stuttgart.de

Uwe-Philipp Kaeppeler and Paul Levi are with the Institute of Parallel and Distributed Systems, Image Understanding Department, University of Stuttgart, Germany, {firstname.lastname}@ipvs.uni-stuttgart.de

same time raises important social, legal and ethical questions – especially if persons are captured by the camera. This raises a series of important questions, which have to be addressed when designing an image based sensor for context acquisition. Thus the remainder of our paper is structured as follows. In Section II important related work is summarized thereby setting the field for our own work. Section III presents a basic framework for context exploitation from image data for non-expert users. Section IV discusses important social, legal and ethical issues and points out further requirements for system design. In Section V a first prototype is presented and evaluated. Finally, we point out future research topics and challenges in Section VI and conclude our work in Section VII.

II. Related Work

Providing sensor information to an infrastructure for context-aware systems based on spatial models has been already investigated by a number of researchers, e.g. Bauer et al. [10] and – building on that – Kaeppeler et al. [11]. Kaeppeler's *SensorContextServer* is a small Debian Linux operated NSLU2 (Network Storage Link for USB 2.0) with different sensors. The SensorContextServer contains sensors for temperature, pressure, humidity, and luminance thereby acting as a small web server speaking AWML (Augmented World Markup Language), an XML-based standardized interchange data format for context-aware applications, so as to ease interoperability between mobile devices and/or the distributed spatial model infrastructure.

Blessing et al. discussed how to transform text-information, e.g. from the internet, to context information, in order to update a spatial model through natural language processing [6]. Blessing et al. called their approach metaphorically a *text sensor* for a spatial model and showed prototypically how to extract and provide traffic report information from the internet to spatial models.

We would like to pick up this metaphoric speech and explore in the following how to build an *image sensor* or *image based sensor*. The required image recognition methods and feature extraction techniques are a vastly researched area. However, as an investigation of existing recognition systems reveals, most of the systems are subject to limitations concerning the preparation of training data or contain restrictions of environmental conditions like illumination and occlusion. Robust, application independent image feature extraction algorithms have been developed by D.G. Lowe [12] (SIFT features), P. Viola/M. Jones [13] (rectangle features) and Bay et al. [14] (SURF features). However, classifiers using these features still need to be adapted to the specific recognition tasks. Although image recognition systems are subject to certain restrictions, application independent image recognition has been a major research area. The goal is the usage of a variety of feature descriptors and the automated selection of the most discriminative ones. This is realized by combining classifiers for different feature descriptors [15, 16, 17, 18]. Opelt et al. [19] discussed automatic feature selection via boosting for solving a variety of image recognition tasks. Although these approaches provide the possibility of learning descriptive models automatically for varying recognition tasks, they rely on the existence of labeled training data. The acquisition of this training data and the manual labeling presents a problem which has not been discussed in depth in connection with this research area. A few systems and applications have been developed which try to build a database of labeled images or objects in images, e.g. using a web-interface [20], a game [21] or specialized software [22]. All of these systems have the same drawback, i.e. they try to motivate the computer vision community to create a large database of labeled images. However, huge effort needs to be put into this creation and the final database will not suit any possible purpose. Heidemann et al. proposed a system based on Self-Organizing Maps (SOM), which allows the labeling of images by arranging them according to their similarity [23, 24]. The proposed system greatly reduces the effort of labeling large amounts of data. A comparable system was proposed in [25] for a visual analytics purpose, where a Self-Organizing Map is used to arrange financial time series data. Neither of both approaches dealt with the requirements introduced by event or situation recognition. Both approaches focused on two-dimensional data only. This may, however, not always be sufficient.

Last but not least, the usage of cameras for context acquisition raises, besides the technical, also a lot of legal, social, and ethical questions, especially when people are involved. As already indicated, image sensors might be deployed especially in environments, such as manufacturing, where the proliferation of sensors or sensor nets is not possible due to the extreme physical conditions. Privacy at the workplace, especially in a so-called "Smart Factory" with Pervasive Computing and Mobile Computing devices acting on a common spatial model infrastructure, has been already explored by Lucke et al. [26] and Wieland et al. [27]. There, the importance of privacy at the workplace is discussed and stressed. However, it is also pointed out, that privacy at the workplace is not free from restrictions. The level of privacy for the worker is the result of a complex negotiation process in which different interests of different parties are balanced or weighed against each other: the employer, the employee, the colleagues of the employee, and third parties such as customers or the public in general.

III. A Basic Framework for Context Exploitation from Image Data

Until today, image data has not been widely used as a source of context. However, the range of possible context information which can be derived from a single image data source is very large. The reason for the restrained usage of image recognition for context exploitation is, that most current image recognition systems are designed for special applications.

The creation of an image recognition system for an

arbitrary task, such as the detection of objects or the evaluation of single statements (like “the window is open”), requires a high level of expert knowledge. Additionally, the creation of such a system comprises the training with a large amount of positive and negative image samples and the category labels for this image data.

The goal of the proposed framework is to facilitate the creation of image based sensors, where one sensor is responsible for the extraction of one context information only. If we want to create a series of these image sensors, two conditions need to be fulfilled: The level of expert knowledge needs to be reduced dramatically and the tasks which need to be performed by the user, i.e. collecting and labeling the training image data, need to be simplified.

Our proposed framework solves these problems by providing a system which assumes no previous knowledge about image recognition tasks and reduces the necessary effort to an acceptable level. This is achieved by an automated selection process of discriminative features and user interfaces for a simplified interaction. The aim of this framework is clearly to reduce the amount of time, necessary for the creation of an image sensor, and to eliminate requirements about the level of knowledge. These conditions render a successful deployment of a network of image sensors possible.

The idea of the framework is to make no assumptions about the upcoming recognition task. Additionally, the framework expects no prior knowledge about pattern recognition from the user. To achieve this completely application independent recognition, we decided to extract a large set of different features from the image data.

Some of these features may not be descriptive, others, however, might be very useful. The features which are useful are automatically selected by the framework in the classification stage. Consider a simple scenario where open windows are to be detected. The training labels indicate those images which show an open window. According to these labels, we can train a classifier for each feature descriptor and combine their results in order to achieve the best classification results on a test set. The combination of these classifiers is realized as a weighted sum of all contributing classifiers. We consider a classifier for a certain feature descriptor as contributing if it decreases the total classification error by at least 5%. We found that a lower improvement rate did not justify the additional computational effort.

The framework is designed as a three-stage architecture, where each stage includes a set of modules, as shown in Fig. 1. The modules included in each stage are not restricted and additional modules can easily be added. The first stage is the region detection. A series of image regions can be detected according to homogeneity, texture or other criteria. Additionally, interest points may be detected. The result of this stage is a spatial description of an image region. This description may also be three-dimensional, i.e. tracking a moving object through a scene results in a spatial description of image regions over time and is treated in the same way as a static image region. The second stage contains feature extraction modules. These modules are applied to all results from the previous stage, that they are able to process, since feature extraction may be able to process only interest points or some regions. The classification stage trains an individual classifier for each feature type. By automatically combining these classifiers, the user is relieved from the task of selecting suitable features, which would require knowledge about the feature extraction modules.

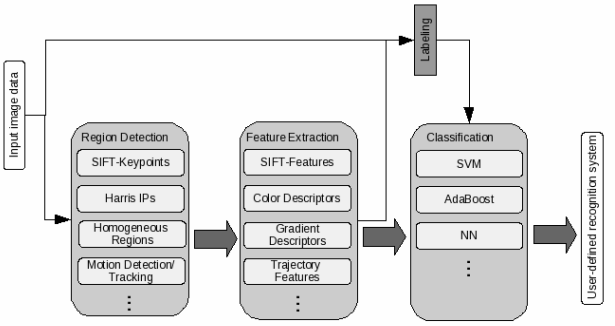

Fig. 1 Three-stage architecture of the proposed framework. The first stage performs the region detection, the second stage extracts features from the detected regions and the third stage provides classifiers for the different feature descriptors. The result of the classification stage is a combination of all individual classification results.

In order to create image based-sensors with this framework, the reduction of the necessary user interaction in the creation of an image sensor is extremely important. Without a major simplification, it is not likely that a series of image sensors will be created. We found the image data acquisition and the training data labeling to be the tasks which require a lot of user interaction. The work for the image data acquisition can be easily reduced by using an automated camera setup which captures images at a certain time, a sequence of images, or a video stream. The proposed framework already includes functionality for the user to easily access and select available data sources. The image data from this source is afterwards processed automatically.

The remaining, time consuming task is the manual labeling of the training image data. Although this task is a major problem in the creation of image recognition systems, it has not been a major area of research. While the arrangement of images according to their similarity is straightforward, the arrangement of time-series data is not. Time-series data is often clustered under the assumption, that the user is interested in unusual events rather than in events that occur often. Since our approach does not comply with this assumption, i.e. the user may not necessarily be interested in unusual events only, we use clustering techniques for time-series data, which identifies similar trajectories. Although clustering techniques, like the Hidden Markov Model based distance by F. Porikli [28], are not perfect for this purpose,

they provide a simplification for the manual labeling task.

IV. Discussion of Ethical and Social Aspects

From an ethical and legal point of view, just monitoring objects and things like in industrial processes does not pose any problems at all. E.g. in manufacturing, the usage of thermographic cameras allows for the remote measurement of temperature of certain processes where the application of sensors or sensor nets is not possible due to the extreme physical conditions. However, if people are involved and monitored, the situation is considered – at least in most European countries – as to be different. If the images/videos a) contain people and b) are of a quality that make it possible to theoretically *identify persons*, the images/videos are subject to data protection law [29, 30, 31]. Their capturing, processing, storage, transfer then needs the informed consent of the people who are subject to the image/video capturing – if no other rights/interests and/or objects of legal protection outbalance/overwrite this basic principle [30, 31]. All in all, from a legal point of view, setting up cameras – if persons are identifiable [sic!] – needs to be a) *required*, b) *appropriate*, and c) *adequate* [30, 31]. Required means that there must be a justifiable reason, a concrete end for which cameras should be installed, i.e. you cannot install cameras just for fun. Appropriate means that setting up cameras is also really necessary and/or that cameras are the right means to reach the given end. Finally, adequate means that setting up cameras is not overdone and also just, reasonable and acceptable for the persons subject to the image/video capturing measures. E.g. in Germany, video surveillance at the workplace is restricted to special cases only. There must be reasonable suspicion documented by facts that an employee acts illegally or contrary to the contract of employment (such as stealing things) [31]. Furthermore, the work council has to be informed, consulted, and asked for approval about the planned surveillance measures. The video surveillance measure mustn't be taken if there are other, less invasive means to prove the guilt of the person in question (thereby protecting, e.g., the rights of other employees).

But what does this all mean for our development of an image based sensor? Mainly it means, that the praised high flexibility of an image based sensor breaks – from the point of view of data privacy – the basic principle of data minimization and data avoidance. The image based sensor is "greedy" capturing much more than just the relevant data needed to reach a certain end. E.g. an image based sensor could be used to detect open windows in a room. At the same time, the sensor also records information about what is going on outside (such as weather, traffic, behavior of people) and what happens inside the room, including the people working there. Therefore, the system design of our image based sensor has to be adjusted accordingly, so that it avoids the above pitfalls and acts as a kind of virtual sensor. First and foremost this implies designing the image sensor as an integrated, closed system, a system, which does not a) forward images or streamed videos, but only measured values like a real sensor (such as temperature values or values which window is open or not) and which b) does also not store or record images/videos locally. This system design would comply with the basic principles of purpose specification/limitation and data minimization/avoidance thus creating a system which could be considered from an ethical and legal point of view as acceptable and therefore also deployable. These restrictions apply to the *operating phase* of the image sensor.

Concerning the *training phase* of the image sensor, it is absolutely necessary to view the image data in order to provide category labels for the training data. However, it is not a valid option to just record everyday life in a factory, since all the data worth protecting would be visible to the person training the system, thus being indeed some kind of systematic surveillance and clustering of the captured images. Therefore, there is no way round explicitly training the system in an artificially created situation by the people who are setting up the system. Some people could now argue, that before releasing the images to the persons training the system, faces could be automatically anonymized or pixelated out. However the problem with that is that a) the anonymization algorithm might miss certain person or b) that even with pixelating out faces too many characteristics remain unobfuscated (such a clothing, posture, shape of the body) thus still allowing for the identification of certain persons [32, 33]. So, in order to be on the safe side, there is no way round explicitly training the system in an artificial situation in which only the training people are visible in the images/videos.

V. Prototypical Implementation and Results

The framework presented in Section III was implemented in MATLAB, in order to exploit the variety of existing detection and feature extraction algorithms.

A prototypical implementation of an image sensor is a door sensor, which continuously tracks the state of the door, i.e. opened or closed. The training data of this sensor is very critical, since the knowledge about persons walking in and out of a door may lead to conclusions about the behavior of these persons, i.e. the time the person does not spend at his or her workplace. Aiming at creating an integrated/closed system, i.e. a sensor which only outputs values and no image data, the door image sensor only returns a Boolean value, which indicates whether the door is open (true), or closed (false).

The system setup consists of a stationary IP camera with restricted access for authorized users only. The creation of the image sensor, including the training, was performed on a workplace computer with Internet access. In addition, the workplace computer was used to apply the sensor in operation mode. The purpose of the image sensor is to update the state of the door in the NEXUS *spatial model* via XML communication.

Since the proposed framework uses state of the art algorithms for the region detection, the feature extraction and the classification stage, we are not able to report new results.

The contribution of this framework is the possibility of creating image sensors with reasonable effort.

In order to discuss the consequences of the requirements implied by ethical and legal issues, i.e. the usage of synthetic training data, we created two image sensors: The first one is trained on synthetic training data, i.e. images acquired in a fake scenario with instructed participants, and the second one is trained on real image data. Fig. 2 shows an exemplary set of training images for both data sets.

The training data for the synthetic image sensor consisted of a total of 213 training images taken with door closed or opened at varying angles. The lighting conditions were changed manually during the data acquisition by dimming the light. However, the changes of the lighting conditions are barely perceptible, see Fig. 2 (images f-j). The real training data consisted of 266 frames, acquired in the course of a day. Frames were captured as soon as a major difference was detected, i.e. the door was opened, and 10 frames of the closed door were captured afterwards.

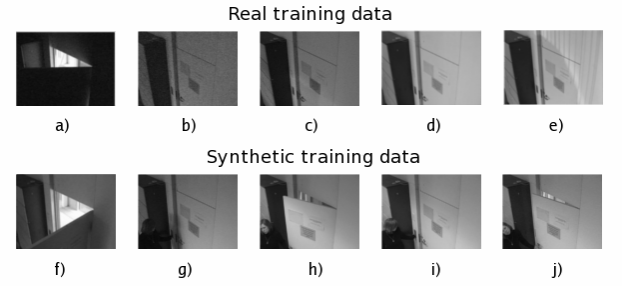

Fig. 2 Exemplary training images from the real (a-e) and synthetic (f-j) training images data set. The images taken from the real training data set display a significant change in lighting conditions.

Both sensors resulted in a combination of rectangle features and color features, with minor differences in the weighting. The test data was sampled over 24 hours every 6 seconds. Frames showing an opened door made up approximately 30% of the test data set. The sensor trained on synthetic data yielded a correct classification rate of 70.4%, whereas the sensor trained on real data yielded a classification rate of 95.8%. A summary of the results is shown in Table I.

An investigation of the misclassified frames in the synthetic scenario indicates that the majority is caused by changes in the lighting conditions which were not covered by the restricted training data set. The lighting changes caused by the daylight have much more effect on the global illumination than the headlights in the room, which were turned on and off during the acquisition of the synthetic training data. Additionally, artificial edges caused by shadows, as can be seen in Figure 2e), may lead to wrong conclusions if they are not sufficiently represented in the training data set.

Although this scenario is very simple, it does reflect a major problem in computer vision and recognition systems. Of course it is possible to acquire a synthetic training data set with a larger variety in lighting conditions. However, the effort and time necessary for this acquisition presents a major obstacle for the creation of image sensors on a large scale.

TABLE I
SUMMARY OF THE TWO COMPARED SENSORS

Sensor trained with	synthetic data	real data
Training data set	213 frames	266 frames
Test data set	13140 frames	
Correct classification rate	70.4%	95.8%

VI. FUTURE WORK: TOWARDS A CLOSED SYSTEM

The discussion has a major implication on the design of the system. The training phase must be performed on a synthetic data set and the operation phase must ensure the privacy of persons captured by the image data. In operation mode, the system must not save the image data for a longer period than necessary for the processing. This issue is important for legal and ethical reasons and can therefore not be realized as an option which may be bypassed. Instead, the framework must ensure that image sensors do not save image data permanently.

In order to implement this restriction we need to focus on two cases, temporary data storage and long-term data storage. Temporary data storage may be necessary inside region detection modules when changes between consecutive frames are of interest. This temporary storage happens inside the module and the data is deleted as soon as it is no longer necessary. To prevent unauthorized access to the temporary data, the system needs to be deployed as a closed system. In the second case, image data is written to the hard drive. This may be necessary if the amount of image data is too large and cannot be processed at once.

While the necessity of temporary storage is indisputable, the long-term storage of data needs to be prevented. This can be achieved by making the system self-aware of its current task. While the system is free to save image data in the training mode, the image sensor itself is subject to certain restrictions, which comprise the prohibition of long-term data storage. Creating this self-awareness by giving the system a training and operation mode is a necessary measure. However, the major drawback of this mechanism is that it is not visible from the outside, whether long-term data storage is in fact prevented.

In future we plan to move our image sensors to a small Debian Linux operated NSLU2 as used in [11] and connect a simple small webcam to it thus creating some kind of real “integrated, closed system.” This small system is able to handle external storage like flash drives or hard disks connected via USB. An implementation of a server that is accessed via network could provide the symbolic measurement results of our image sensor but prevent the access to the images of the camera. An internal 8MB flash, which is not accessible via network, can be used to store calibration datasets for the image processing. The big advantage of such a system is the possibility of removing the memory which is necessary for permanent storage of the

image data. We suggest the development of an integrated, closed system consisting of a camera and an embedded system similar to the NSLU2, where the USB plug is mounted next to the camera lens. Therefore it would be possible for the person in front of the camera to see if any storage device is connected and whether the device is technically capable of storing images at the moment or not.

As stated in Section IV in some countries any capturing, processing or storage of data needs the informed consent of persons who could possibly be hereby identified. In cases where e.g. employees are in the range of coverage of an image sensor, the acceptance should be higher if they can easily check for themselves that permanent image storage is physically prevented.

In turn the identical hardware together with a connected USB storage could be used to operate our image sensor in training mode, whereas the calibration of the sensor on real training data leads to better classification rates as has been proved by our prototypical implementation.

The actual creation and deployment of this integrated, closed system is, however, not a scientific challenge.

VII. Conclusion

In this paper, we presented a framework for the exploitation of context information from image data. The major contributions of this framework are its extensibility and its usability. The framework is designed for users with no prior knowledge about image recognition techniques. The proposed framework facilitates the creation of image recognition systems, i.e. image sensors, tremendously. Moreover, the discussion of different social, legal, and ethical issues showed that – if people are involved – the above framework needs to be implemented as an integrated, closed system so as to be considered as acceptable and thus also as deployable. A prototypical implementation and first test have shown that the proposed system is effective and efficient. However, it still remains an open question how the image sensor is perceived by e.g. the people at work. For them they only see a camera. While some won't care about it, others might feel observed, monitored and tracked and won't trust that the camera is programmed in the right way (i.e. capturing only symbolic sensor information and not streaming or storing images, videos or information about their behavior or movements).

All in all, setting up cameras is a delicate issue which needs more research especially in conjunction with legal scholars.

A topic of future research will not only be the exploitation of context information from image data and the updating of the spatial model, but also the evaluation of the correctness of already modeled information and the consistency of the spatial model at all. We therefore plan to combine our *image sensor* with the already developed *text sensor* and other real, physical sensors as are already available with the *SensorContextServer*. For checking consistency we furthermore want to concentrate on image and video data already provided by different internet communities. E.g., there is a growing number of freely available web cams showing the traffic volume and road-conditions. By extracting information with an image sensor from these web cams and combining/checking them against information acquired with the text sensor about the traffic volume, we want to make information in the spatial model more exact and reliable thus allowing in future for more robust context-aware systems.

Acknowledgements

This work was developed within the Nexus project (collaborative research centre/SFB 627), which is supported by the German Research Foundation (DFG).

References

[1] A. K. Dey, G. D. Abowd, P. J. Brown, N. Davies, M. Smith, and P. Steggles. Towards a better understanding of context and contextawareness. In Proceedings of the 1st International Symposium on Handheld and Ubiquitous Computing, 1999.

[2] M. Großmann, M. Bauer, N. Hoenle, U.-P. Kaeppeler, D. Nicklas, and T. Schwarz. Efficiently managing context information for large-scale scenarios. In Proceedings of the 3rd IEEE International Conference on Pervasive Computing and Communications, 2005.

[3] K. Henricksen and J. Indulska. A software engineering framework for context-aware pervasive computing. In Proceedings of the 2nd IEEE International Conference on Pervasive Computing and Communications, 2004.

[4] M. Roman and R. H. Campbell. Gaia: Enabling active spaces. In Proceedings of the 9th ACM SIGOPS European Workshop, Kolding, 2000.

[5] D. Salber, A., K. Dey, and G. D. Abowd. The Context Toolkit: Aiding the development of context-enabled applications. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: The CHI is the Limit, 1999.

[6] A. Blessing, A.; S. Klatt, D. Nicklas, S. Volz, and H. Schuetze. Language-Derived Information and Context Models. In Proceedings of the 3rd Workshop on Context Modelling and Reasoning at 4th IEEE International Conference on Pervasive Computing and Communication perCom, 2006.

[7] U.-P. Kaeppeler, R. Benkmann, O. Zweigle, R. Lafrenz, P. Levi: Resolving Inconsistencies in Shared Context Models using Multiagent Systems. In R. Dillmann and W. Burgard (eds.): Proceedings of the 10th International Conference on Intelligent Autonomous Systems: IAS-10; Baden Baden, Germany, July 23-25, 2008.

[8] M. Bauer, L. Jendoubi, and O. Siemoneit. Smart Factory – Mobile Computing in Production Environments. In Proceedings of the MobiSys Workshop on Applications of Mobile Embedded Systems WAMES, 2004.

[9] L. M. Hilty, C. Som, and A. Koehler. Impacts of Future Information and Communication Technologies on Society and Environment. In G. Banse, I. Hronszky, and G. Nelson (eds.): Rationality in an uncertain world. edition sigma, 2005, 205-290.

[10] M. Bauer, C. Becker, J. Haehner, and G. Schiele. ContextCube – Providing context information ubiquitously. In Proceedings of the 3rd International Workshop on Smart Appliances and Wearable Computing, 2003.

[11] U.-P. Kaeppeler, A. Gerhardt, C. Schieberle, M. Wiselka, K. Haeussermann, O. Zweigle, and P. Levi. Reliable situation recognition based on noise levels. In Proceedings of the 1st International Conference on Disaster Management and Human Health Risk, 2009. (Forthcoming)

[12] D. G. Lowe. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, 60:91–110, 2004.

[13] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. pages 511–518, 2001.

[14] H. Bay, A. Ess, T. Tuytelaars, and L. van Gool. Speeded-up robust features (surf)). Computer Vision and Image Understanding, 110(3):346 – 359, 2008. Similarity Matching in Computer Vision and Multimedia.

[15] J.Kittler, M. Hatef, R. P.W. Duin, and J. Matas. On combining classifiers. IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(3):226–239, 1998.
[16] Z. Stejic, Y. Takama, and K. Hirota. Mathematical aggregation operators in image retrieval: effect on retrieval performance and role in relevance feedback. Signal Processing, 85(2):297 – 324, 2005. SI on Content Based Image and Video Retrieval.
[17] M. Varma and D. Ray. Learning the discriminative power-invariance trade-off. In Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on, pages 1–8, Oct. 2007.
[18] O. R. Terrades, E. Valveny, and S.Tabbone. Optimal classifier fusion in a non-bayesian probabilistic framework. IEEE Trans. Pattern Anal. Mach. Intell., 31(9):1630–1644, 2009.
[19] A. Opelt, A. Pinz, M. Fussenegger, and P. Auer. Generic object recognition with boosting. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 28(3):416–431, March 2006.
[20] B.C. Russell, A. Torralba, K.P. Murphy, and W.T. Freeman. LabelMe: A Database and Web-based Tool for Image Annotation. Int. J. Comput. Vision, 77(1-3):157-173, 2008.
[21] L. von Ahn and L. Dabbish. Labeling images with a computer game. In Human factors in computing systems. CHI 04. SIGHI conference on. Pages 319-326, 2004.
[22] S. Ayache and G. Quenot. Trecvid 2007 collaborative annotation using active learning. In In Proceedings of the TRECVID 2007 Workshop, 2007.
[23] G. Heidemann, A. Saalbach, and H. Ritter. Semi-automatic acquisition and labeling of image data using SOMs. In ESANN, pages 503–508, 2003.
[24] H. Bekel, G. Heidemann, and H. Ritter. Interactive image data labeling using self-organizing maps in an augmented reality scenario. Neural Netw., 18(5-6):566–574, 2005.
[25] T. Schreck, J. Bernard, T. von Landesberger, and J. Kohlhammer. Visual cluster analysis of trajectory data with interactive kohonen maps. Information Visualization, 8(1):14–29.
[26] D. Lucke, E. Westkaemper, M. Eissele, T. Ertl, O. Siemoneit, and C. Hubig. Privacy-Preserving Self-Localization Techniques in Next Generation Manufacturing. An Interdisciplinary View on the Vision and Implementation of Smart Factories. In Proceeding of the 10th International Conference on Control, Automation, Robotics and Vision ICARCV, 2008.
[27] M. Wieland,C. Laengerer, F. Leymann, O. Siemoneit, and C. Hubig. Methods for Conserving Privacy in Workflow Controlled Smart Environments. A Technical and Philosophical Enquiry into Human-Oriented System Design of Ubiquitous Work Environments. In Proceedings of the 3rd International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies Ubicomm, 2009. (Forthcoming)
[28] F. Porikli. Trajectory distance metric using hidden markov model based representation. In Proceedings of the 8th IEEE European Conference on Computer Vision, PETS Workshop, 2004.
[29] Unabhaengiges Landeszentrum fuer Datenschutz Schleswig-Holstein. Videoueberwachung und Webkameras (German). Blaue Reihe Vol. 4. URL=https://www.datenschutzzentrum.de/blauereihe/blauereihe-video.pdf
[30] M. Lang. Private Videoueberwachung im oeffentlichen Raum. Eine Untersuchung der Zulaessigkeit des privaten Einsatzes von Videotechnik und der Notwendigkeit von § 6b BDSG als spezielle rechtliche Regelung (German). Hamburg, 2008.
[31] A. Rossnagel, S. Jandt, J. Mueller, A. Gutscher, and J. Heesen. Datenschutzfragen mobiler kontextbezogener Systeme (German). Wiesbaden, 2006.
[32] O. Siemoneit, C. Hubig, M. Kada, M. Peter, and D. Fritsch. Google Street View and Privacy. Some thoughts from a philosophical and engineering point of view. In Proceedings of the 5th Asiac-Pacific Conference on Computing and Philosophy, 2009.
[33] M. Kada, M. Peter, D. Fritsch, O. Siemoneit, and C. Hubig. Privacy-Enabling Abstraction and Obfuscation Techniques for 3D City Models. In Proceeding of the 2nd SIGSPATIAL ACM GIS International Workshop on Security and Privacy in GIS and LBS, 2009. (Forthcoming)