

TRUSS: Tracking Risk with Ubiquitous Smart Sensing

Brian Mayton, Gershon Dublon, Sebastian Palacios, and Joseph A. Paradiso
Responsive Environments Group, MIT Media Lab, Cambridge, MA 02139
Email: {bmayton, gershon, joep}@media.mit.edu, sebastian.palacios@gatech.edu

Abstract—We present TRUSS, or Tracking Risk with Ubiquitous Smart Sensing, a novel system that infers and renders safety context on construction sites by fusing data from wearable devices, distributed sensing infrastructure, and video. Wearables stream real-time levels of dangerous gases, dust, noise, light quality, altitude, and motion to base stations that synchronize the mobile devices, monitor the environment, and capture video. At the same time, low-power video collection and processing nodes track the workers as they move through the view of the cameras, identifying the tracks using information from the sensors. These processes together connect the context-mining wearable sensors to the video; information derived from the sensor data is used to highlight salient elements in the video stream. The augmented stream in turn provides users with better understanding of real-time risks, and supports informed decision-making. We tested our system in an initial deployment on an active construction site.

I. INTRODUCTION

Accidents on construction sites are extremely costly, result in lost time, and are sometimes tragic. As such, construction companies are interested in developing technologies that prevent accidents before they happen. Construction workers engage in various activities that may expose them to serious hazards, such as falling from rooftops and scaffolding, unguarded machinery, being struck by heavy construction equipment, electrocutions, and dangerous dust particles. The Occupational Safety and Health Administration (OSHA) provides a guide for the abatement of physical hazards in the construction industry with suggestions for eliminating, controlling or mitigating them [1]. In reality, no foreseeable technology can completely replace expert human supervision, and accidents happen despite strict adherence to protocol. At the same time, even the best safety supervisors cannot possibly observe an entire construction site at once; standard practice prescribes intermittent inspection, typically conducted in scheduled walk-throughs, and attention is spread across many activities. Sensor networks can extend the sensory reach of the safety supervisor in space and time, allowing continuous risk assessment both on- and off-site, while directing the supervisor's attention to those situations that most demand it. In this paper, we present TRUSS, or Tracking Risk with Ubiquitous Smart Sensing, a system that infers and renders safety context on construction sites by fusing data from wearable devices, a flexible distributed sensing infrastructure, and video. In our system, expert, human-in-the-loop engagement is not replaced by pure automation, but supported by smart sensing and visualization.

Our system fuses multimodal data streams from infrastructure and wearable sensors to support a visual analytic [2] user interface that provides safety supervisors with context they can use to evaluate evolving risks. Wearable sensors stream real-time levels of dangerous gases, dust, noise, light quality, precise altitude, and motion to base stations that synchronize the mobile devices, monitor the environment, and capture video. At the same time, low-power video nodes track the workers as they move in and out of the field of view, attempting to re-identify the tracks using information from the sensors. These processes together connect the context-mining wearable sensors to the video; information derived from the sensor data is used to highlight salient elements in the video stream, and the augmented stream in turn provides users with better understanding of real-time risks for decision support. For example, altitude signals

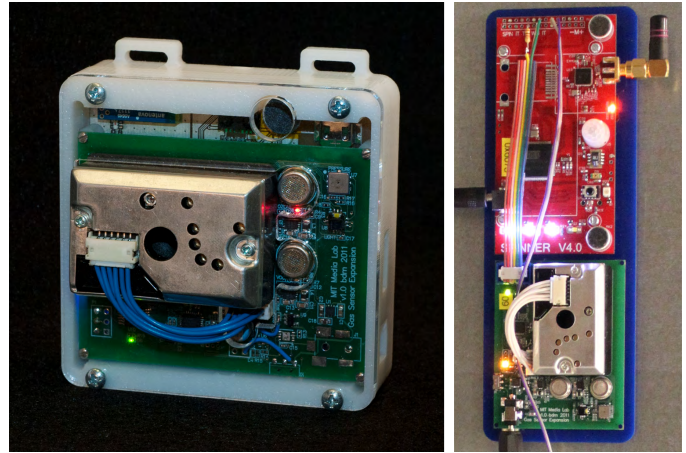


Fig. 1. Wearable sensor badge (left) and base station (right) containing environmental, inertial, and altitude sensors, as well as an 802.15.4 radio for real-time data streaming.

from the wearables are fused with track signals from the cameras to infer consistent labels for workers as they move in and out of the camera views, providing a means to visualize individual risks using augmented video (e.g. exceeding an altitude threshold or working below others). The sensor fusion also allows us to use relatively simple vision trackers that can run on low-power devices onsite.

We tested TRUSS on an active construction site, instrumenting 3 steelworkers for a period of 2 weeks, during which they erected several large steel structures in a new building. Accidents are fortunately rare and we did not observe one during our study, but both the steelworkers and the supervisors viewed TRUSS as a valuable tool that would improve their safety if deployed as a real-time application. The data collection was conducted to support a proof-of-concept, with the goal of real-time operation in the future.

II. RELATED WORK

A number of researchers have applied wireless sensor networks (WSNs) and wearable electronics to improving worker safety. Niu, et al. develop a system for wirelessly monitoring methane concentrations in mines, focusing on intelligent sampling to reduce traffic [3]. Angove and O'Flynn present their Bluetooth air-quality monitor in [4]; our environmental monitoring hardware inherits from their work. Fugini, et al. theorize wearables that would integrate seamlessly with services to warn their wearers of risks, raise alarms for others, or connect workers to emergency personnel when appropriate [5]. Other researchers have focused on algorithmic activity recognition using wearables. Lukowicz, et al. recognize workshop tasks like sawing, hammering, and turning a screw, using microphones and accelerometers mounted on a user's arm and training hidden Markov models (HMM) for classification [6], [7].

Camera and wearable sensor fusion is motivated by the premise that significant algorithmic challenges and computational costs of parsing video can be mitigated by the labeling of tracked objects through fusion of ID-linked sensor data with synchronized video.

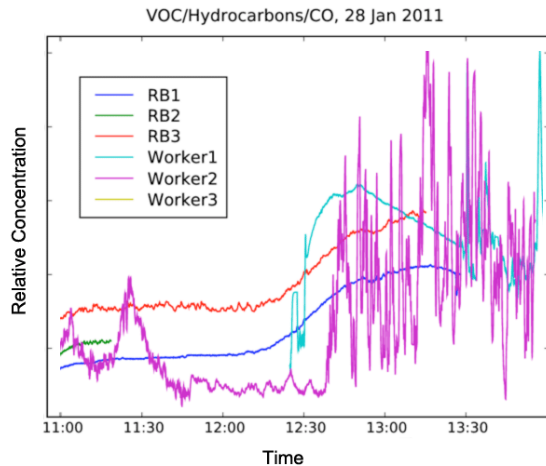


Fig. 2. Volatile gas concentrations from wearable and infrastructure sensors over a 3-hour period on the worksite. Painting begins halfway into the plot; fumes are detected by both base stations (red and blue) and wearables (cyan and magenta.) Worker 2's (magenta) measured levels are fluctuating as he moves between areas close to the painting and areas with better ventilation.

Candidates for this fusion include inertial sensors like accelerometers, or other sensors that can capture some shared state also available to a camera. Teixeira, et al. show in [8] that people wearing accelerometers can, under certain conditions, be reliably linked to their corresponding camera tracks by matching gait timing parameters between the accelerometers and the camera. However, this strategy fails when subjects are not walking, and when camera positions are not ideal. In [9], the authors generalize this idea, forming an HMM for each person consisting of a measurement from the inertial sensor and a measurement from the camera and applying maximum *a posteriori* estimation to generate the most likely matches.

A number of researchers encode information in images conveyed to users through interactive interfaces. Teodosio and Bender define a class of images called Salient Stills which capture and convey information from across times in single, still images [10]. Our work in TRUSS builds on this idea, using data from sensors in the scene to select regions of video for extraction.

Finally, there have been a number of recent efforts to create sensor network user interfaces that support open-ended exploration of multimodal data. Several location- and orientation-aware handheld devices allow their users to move through the network, and provide interfaces for defining inference rules [11]. DoppelLab is a 3-d virtual environment for exploring multimodal sensor data and its relationship to architectural space [12].

III. THE TRUSS SYSTEM

The TRUSS system is composed of three main hardware components: battery-powered wearable sensor devices, externally powered base stations, and cameras attached to embedded Intel Atom computers. The components are time-synchronized using the ZigBee radios, enabling applications that fuse data from the independent sources.

The embedded computer acts as a network bridge between the low-power ZigBee network and a fixed-infrastructure WiFi or GSM network, synchronizing the ZigBee network clocks and enabling remote connections. The computer also runs a video subsystem with cameras, a computer vision library, a video encoder, and a streaming server. A more powerful remote server could perform further operations on the video before it reaches users, though this is not the case in the current system.

The fixed base stations host a set of sensors and connect to a daughterboard designed for environmental monitoring. The onboard sensors include a PIR motion sensor that can be used to trigger data collection or processing when workers are detected. Microphones pick up loud crashes, machine noise, and yelling. An infrared "sociometric" sensor [13] on both the base stations and mobile nodes is used to detect where workers are facing, and when and how they work together. Light level and color sensors keep track of lighting conditions to detect welding or anomalous flashes of light. Barometric pressure sensors on the base stations, in tandem with pressure sensors on each wearable, can provide relative altitude between the base stations and wearables with sub-meter precision [14]. Both base stations and wearables carry the environmental monitoring system, with sensors that measure levels of particulates and uncalibrated levels of volatile organic gases (VOC), hydrocarbons, and ozone. Magnetic strips allow the base stations to be rapidly deployed and moved as necessary by the workers on a construction site with many stationary ferrous surfaces, such as scaffolding and machinery.

The wearable node includes a single microphone, light level and color sensors, the IR sociometric sensor, and a temperature sensor. It also carries a 3-axis accelerometer, 3-axis gyroscope, and 3-axis magnetometer. ZigBee radios on both the mobile and infrastructure nodes are used for synchronization and can be used for rough localization of the mobile devices to within an area of 5-10 meters radius using radio link quality (due to hardware issues, we were unable to monitor the much more effective radio signal strength indicator). In our first test, the radio-location accuracy suffered from a very challenging environment, facing large metal obstructions and too low a density of base stations to provide useful location information.

We explored several options for mounting the wearable sensor node on the workers, ultimately offering them the options of belt-worn and lanyard-worn nodes. Helmet mounting was also considered, but our study subjects were to be frequently switching between welding masks and standard construction hard hats.

The gas sensors we included were chosen for the specific construction site targeted for our first deployment, which was to involve steelworkers who would be working at height while cutting and welding steel. These activities increase the risk of falling objects and generate flammable or otherwise harmful gases. Elevated ozone levels indicate arc welding activity, while VOCs and hydrocarbons indicate flammable solvents. Spatiotemporally proximate elevated levels of both would present a risk of ignition.

A. Tracking and Sensor Fusion

The cluttered and constantly changing environment of a construction site creates a particularly challenging tracking problem. Frequent occlusions, reflections and flashes from welding, and a constantly changing background rules out pure computer vision, especially as workers move between discontinuous camera views. Radio-based tracking alone is also not realistic, as base stations are moved often and large metal objects are in constant motion. Inertial dead-reckoning is prone to accumulating error. Cameras are quite effective as long as the worker does not leave the field of view [15]; the camera alone cannot recover a lost track.

We opted for a relatively simple vision pipeline: a mean-shift blob tracker [16] operates on the image after a process of frame-differencing, thresholding, morphological operations, and a weighted moving average to smooth the motion. This approach works well to find and track people who move with some frequency, but also finds anything else (machinery, etc) that moves, though this problem can be mitigated slightly by a well-chosen search radius and blob size

threshold. Still, the tracker can not segment multiple people when they occlude each other, causing problems when they separate again and the tracker cannot resolve the path ambiguity.

The cluttered environment of our target construction site was not conducive to a gait-based camera-wearable fusion technique like [8], as there was little space to walk freely, many occlusions, and long periods of relatively stationary activities (e.g. welding in a cluttered area). However, in the environment of our test and in many similar scenarios, workers are often ascending and descending on lifts and ladders in cramped spaces. A fusion between the pressure-based altitude sensor and the camera tracker can recover worker ID after track ambiguities. This approach fails if the workers are close to each other at the same altitude (working on the same lift), but the system can subsequently recover. In this scheme, the tracker fuses three pieces of information: the number of workers in the field of view, the altitudes of each worker provided by their wearable pressure sensors, and the altitudes inferred from each tracked blob; a simple flowchart-based algorithm performs logical and nearest-neighbor operations on the track and altitude signals. This approach is intended as a test of the data correspondence in the prototype system. Under the conditions described above, the system can reject spurious tracks caused by moving equipment and shadows, as well as identify multiple workers in a scene after the tracker's state has been cleared.

IV. SAFETY CONTROL PANEL

The TRUSS Control Panel is a software user interface for safety managers that combines data streams from multiple sensors to better communicate the changing context around each worker. The interface augments video of workers on the construction site with information from their wearable sensor nodes, and allows users to set thresholds and priorities on single data streams or combinations of streams. This person-centric information architecture imagines a worker safety bubble metaphor, where a sphere of some variable radius encloses a worker's local context, and highlights the intersection of that context with other bubbles centered on workers, machinery, or material.

This approach is intended to keep the user's focus on the workers while highlighting the causes of common accidents. In many cases, coincident activities and circumstances that might individually be safe can together be catastrophic, like a worker passing underneath another carrying heavy objects at height, or flammable gases coming into contact with sparks from a welder. By allowing interactive control over the sensor thresholds, selection criteria, and critical ranges, the information visualization can adapt to changing conditions as the expert user sees fit. In this way, the interface plays a decision-supporting or visual analytic role for an expert user.

Fig. 3 shows the TRUSS user interface to multimodal sensor data and video. The interface is designed to prioritize and highlight what its user deems to be the most relevant information, while at the same time letting all the data through in the background to provide context. On one side of the interface, a stitched-together view from multiple cameras provides coverage of the scene; on the other, a spotlight view highlights areas of interest. In both views, workers are augmented with tinted circles reflecting both the levels (through intensity) and types (through color) of dangerous gases around them. When the spotlights intersect, colors mix, highlighting the extent as well as the nature of the event. Building on the safety bubble metaphor, the visualization treats areas of overlap as particularly worthy of user attention; the radii of the bubbles can be set in the interface. The outline of each circle reflects whether the enclosed worker has exceeded an altitude threshold. The graphs show recent history, which places the instantaneous video augmentation in a temporal context.

The TRUSS interface also offers a mode that composites video from all the cameras into a single stream, weighting each input video by the user's choice of any sensed parameter (or group of parameters). For example, the software might emphasize video of those individuals exposed to the highest levels of ozone, or those most proximate to others working at height. In our first deployment, this feature was not particularly useful because of the relatively small number of workers in the test; further development is planned for larger deployments.

V. USER STUDY

We tested TRUSS with 3 steelworkers over 2 weeks on an active construction site. During this time, the workers erected several large catwalk structures surrounding a set of 2-story air handlers on the penthouse floor of a building under construction. The workers' primary activities were arc welding, cutting steel, and carrying heavy material, all while using ladders, lifts, and cranes for rigging large steel frames and platforms. The study was designed to test our system by collecting data to verify the sensors and hardware, and to support the development of the safety control panel software using real data. We also sought to gauge both the workers' and supervisors' attitudes to safety monitoring, and to make sure the physical system could fit into their existing workflow without causing problems. We distributed the wearables and magnet-backed base stations to workers at the start of each workday, and deployed the video PC and cameras ourselves.

We encountered several problems during the deployment, both systemic and environmental. Embedded software instability caused occasional system hang-ups. In addition, heat from the gas sensors caused errors on the pressure sensor. The environmental challenges included extreme, bit-scrambling electric fields caused by welding, as well as metal filings falling into vents and causing shorts. In order to allow air to flow freely into the gas sensors, the boards were left mostly uncovered, resulting in at least one catastrophic short due to worker sweat (that worker later received a better-sealed device). Taken together, these problems resulted in partial data coverage. Ultimately, we collected enough data to prove concepts and establish correspondence between activities and measurements, but not enough to build robust models. The next revision of the hardware addresses many of these issues, and new deployments are planned.

There was some concern that workers would view the sensors and cameras with suspicion, as tools of surveillance. We made it clear that workers had absolutely no obligation to participate, and could terminate the study at any time. We were surprised to find workers not only agreeable, but extremely supportive of the research. The workers reported concern for their own safety, and a hope that sensor technologies could help them better understand and respond to risks. A construction manager expressed the most concern for this issue, but characterized it as a problem limited to the small number of workers who would already be looking for ways to take unsafe shortcuts. We did not encounter this attitude in the study.

After the deployment, we showed the safety control panel to a third party construction industry expert. He expressed several important concerns and recommendations, as well as interest in further study to be conducted on new sites. He reported that worker behavior and state of mind are major risk factors that we are not presently considering, giving the example of a worker who may be frustrated by heavy traffic on their morning commute, resulting in reckless behavior later in the day. Still, he found our notion of a personal safety bubble compelling. He noted that the interface makes some axes of otherwise invisible context clear to users, and could be useful for training. He also noted that 80% of safety management is prevention and 20% is field control, and expressed that TRUSS could fit well in both. Finally, he asked

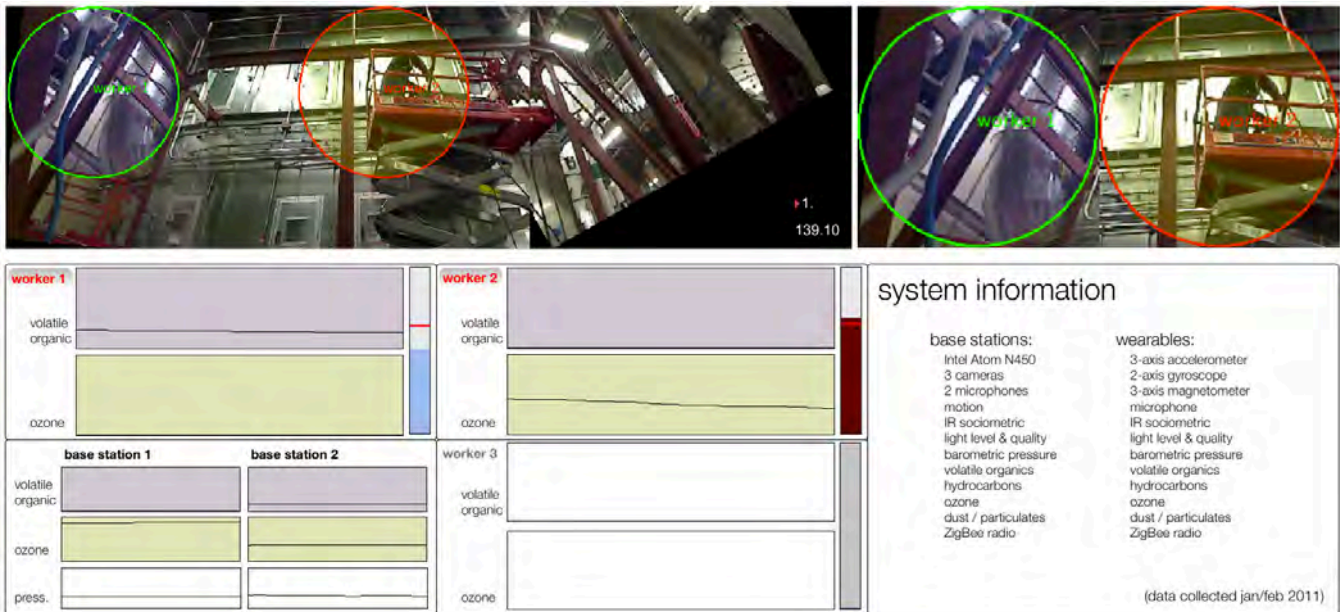


Fig. 3. Safety control panel interface. The elevated level of ozone surrounding the worker on the right indicates that he has been welding, while the elevated level of volatile organic gases around the worker at left indicate an activity involving a paint or aerosol. Overlapping bubbles of these kinds would suggest an increased risk of fire or explosion, and might be cause for alarm.

if there was some way he could see the whole site in a macroscopic way that would expose faults and other points of interest for further exploration through a more detailed interface like TRUSS. Shown our prior work on macroscopic sensor network interfaces from [12], he strongly suggested that the two interfaces be integrated.

VI. CONCLUSION

We presented a system for tracking risks to construction workers using mobile and flexible infrastructure sensing together with cameras. Using data collected from a test of the hardware system, we developed a prototype UI for safety supervisors based on a sensor-supported safety bubble metaphor. A simple but effective sensor fusion algorithm was developed to identify tracked workers in video using wearable barometric pressure sensors. There are clear avenues for applying sensor fusion in future work, between microphones on workers and in the infrastructure, or inertial sensors and cameras, for example. New radio-location strategies could, together with a vision tracker, effectively solve the video correspondence problem outright. We believe that intuitive interfaces backed by intelligent algorithms can better connect safety supervisors to the complex, evolving context of a construction site.

A brief video is at <http://youtu.be/O7Des085tTw>.

ACKNOWLEDGMENTS

This work was supported by Enel S.p.A, the Intel Corporation, and the research sponsors of the MIT Media Lab. The work was part of a larger research initiative at MIT, called Future Enel. Thanks to Stephen Miles, Carlo Ratti, Hiroshi Ishii, and the team.

REFERENCES

- [1] OSHA 1926: Safety and Health Regulations for Construction, 2012.
- [2] J. J. Thomas and K. A. Cook, "A visual analytics agenda," *IEEE Computer Graphics and Applications*, vol. 26, no. 1, pp. 10–13, Jan. 2006.
- [3] X. Niu, X. Huang, Z. Zhao, Y. Zhang, C. Huang, and L. Cui, "The design and evaluation of a wireless sensor network for mine safety monitoring," *Global Telecommunications Conference, 2007. GLOBECOM'07. IEEE*, pp. 1291–1295, 2007.
- [4] P. Angove and B. O'Flynn, "Air-quality Monitoring for Pervasive Health," *IEEE Pervasive Computing*, vol. 9, no. 4, pp. 48–50, 2010.
- [5] M. Fugini, G. M. Conti, F. Rizzo, C. Raibulet, and L. Ubezio, "Wearable services in risk management," *Web Intelligence and Intelligent Agent Technologies, 2009. WI-IAT'09. IEEE/WIC/ACM International Joint Conferences on*, vol. 3, pp. 563–566, 2009.
- [6] P. Lukowicz, J. A. Ward, H. Junker, M. Stäger, G. Tröster, and T. Starner, "Recognizing workshop activity using body worn microphones and accelerometers," *IEEE Pervasive Computing*, 2004.
- [7] J. A. Ward, P. Lukowicz, G. Troster, and T. E. Starner, "Activity Recognition of Assembly Tasks Using Body-Worn Microphones and Accelerometers," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 10, pp. 1553–1567, 2006.
- [8] T. Teixeira, D. Jung, G. Dublon, and A. Savvides, "PEM-ID: Identifying people by gait-matching using cameras and wearable accelerometers," in *IEEE International Conference on Distributed Smart Cameras*, Yale University. IEEE, 2009.
- [9] T. Teixeira, D. Jung, and A. Savvides, "Tasking networked CCTV cameras and mobile phones to identify and localize multiple people," in *Proceedings of the 12th ACM international conference on Ubiquitous computing (Ubicomp '10)*, 2010.
- [10] L. Teodosio and W. Bender, "Salient video stills: Content and context preserved," in *Proceedings of the first ACM international conference on Multimedia*, New York, 1993.
- [11] M. Mittal and J. A. Paradiso, "Ubicorder: A Mobile Device for Situated Interactions with Sensor Networks," *IEEE Sensors Journal*, 2011.
- [12] G. Dublon, L. Pardue, B. Mayton, N. Swartz, N. Joliat, P. Hurst, and J. A. Paradiso, "DoppelLab: Tools for Exploring and Harnessing Multimodal Sensor Network Data," in *IEEE International Conference on Sensors (IEEE Sensors)*. Limerick, Ireland: MIT Media Lab, 2011.
- [13] M. Laibowitz and J. A. Paradiso, "The UBER-Badge, a versatile platform at the juncture between wearable and social computing," *Advances in Pervasive Computing*, 2004.
- [14] W. Brunette, J. Lester, and A. Rea, "Some sensor network elements for ubiquitous computing," *4th International Symposium on Information Processing in Sensor Networks*, 2005.
- [15] T. Teixeira, G. Dublon, and A. Savvides, "A Survey of Human Sensing: Methods for Detecting Presence, Count, Location, Track and Identity," *ACM Computing Surveys*, 2010.
- [16] D. Comaniciu, V. Ramesh, and P. Meer, "Real-time tracking of non-rigid objects using mean shift," in *CVPR 2000*. IEEE Comput. Soc., 2000, pp. 142–149.