There are many ways to capture human gestures. In this paper, consideration is given to an extension to the growing trend to use sensors to capture movements and interpret these as gestures. However, rather than have sensors on people, the focus is on the attachment of sensors (i.e., strain gauges and accelerometers) to the tools that people use. By instrumenting a set of handles, which can be fitted with a variety of effectors (e.g., knives, forks, spoons, screwdrivers, spanners, saws etc.), it is possible to capture the variation in grip force applied to the handle as the tool is used and the movements made using the handle. These data can be sent wirelessly (using Zigbee) to a computer where distinct patterns of movement can be classified. Different approaches to the classification of activity are considered. This provides an approach to combining the use of real tools in physical space with the representation of actions on a computer. This approach could be used to capture actions during manual tasks, say in maintenance work, or to support development of movements, say in rehabilitation.

1. INTRODUCTION

Researchers in the field of Human-Computer Interaction (HCI) have expended a great deal of time and effort on the design and psychology of graphical user interfaces, but significantly less attention to the design of the devices with which users can interact with these interfaces. Our current range of interaction devices restrict people to a very limited number of actions which tend to be performed in series and tend to require only one hand. Many aspects of our everyday lives involve the use of our hands to grasp, manipulate and operate objects in the world around us. We have a well-developed repertoire of movements to allow fine motor control of our hands and fingers, and yet these movements are rarely supported in HCI. More often than not, the flexibility of the human hand is ignored and movements are reduced to either pressing buttons or grasping a mouse to make small, constrained movements in order to control a cursor on the screen.

Developing interactive technology that reflects the richness of dexterous behavior remains a challenge for HCI. There are, of course, exceptions to this statement. Over the past decade or two, pen-based computing has allowed people to hold a stylus and manipulate it much like a pen to write and draw on screen or to hold a stylus to manipulate objects in virtual environments and receive haptic feedback, and in the past few years, gaming devices, such as the Nintendo Wii, have supported hand and arm movements that are similar to those used in sport and dance. One reason for this trend is the desire to produce ‘multi-functional’ devices, such as the mouse, which can be used to perform a variety of functions. These devices share a common underlying approach: the device that is held in the hand is first and foremost intended to be used to act upon virtual objects on the computer screen or in the virtual world. This can be contrasted with the many ways that we use devices (or tools) to act upon real objects in the real world. Not only are the compliances and behaviours of these real objects more complex than their virtual counterparts but also the range of actions we perform with tools to exploit these compliances are more varied. Notwithstanding the fact that a single device is unlikely to satisfy all of the requirements of all users of a computer, there is an obvious irony in the use of the term multifunctional to describe a mouse. The functionality of the mouse lies not in the device (that can only offer the functions of linear movement in the horizontal plane and depression of one, two or three buttons) but in the objects on the graphical user interface that the mouse is used to manipulate. While this provides a means of linking physical activity to a graphical display, in many of our everyday activities, the use of the tool provides implicit feedback to the user. This feedback takes the form of the feel of the tool and the effects that the user can make on objects in the world using the tool. An alternative perspective would be to define the physical device in terms of its functionality, and to
capture user behaviour to manage HCI. This is the approach that underlies Tangible User Interfaces. In broad terms, interaction with tangible user interfaces can be considered as follows: ‘a user manipulates a physical artifact with physical gestures, this is sensed by the system, acted upon, and feedback is given’ [32] p.253). The physical artifacts can range from models of real objects [9] [11][38], to construction blocks [40] to everyday objects that have been adapted to connect to a digital environment, such as the MediaCup [14]. It is this latter class that is the focus of this paper. Thus, one can conclude that interaction devices can be developed as significant components of the computer systems, not only acting as transducers to convert user action to computer response, but communicating all manner of feedback to the user and supporting a greater variety of physical activity.’ [4] p. 276).

The manipulation of pointing devices, such as mouse or joystick or game controllers, requires specific control movements; users can not simply adapt movements that are familiar to them but need to learn new ones. Admittedly, this learning is not particularly onerous because the range of movements permitted is so small. However, it does mean that there is a gap between making a movement in ‘real-life’ and making a movement in order to control virtual objects on a screen. This latter type of movement has the goal of performing actions in order to control something. Some devices, such as stylus and pens, are able to support movements that are similar to learned movements, such as drawing and writing. However, it is interesting to note that there are still some differences between performing these movements with pen and paper versus stylus and screen [4]. This means that the movements that a person is performing can be considered partly ‘natural’ (ie learned and practiced in everyday life) and partly a response to the demands of the computer. Rather than the user having to learn sets of movements that the computer can interpret, the computer could be made to adapt to the sets of movements that the person naturally wants to make. By presenting people with a well-defined task, such as hitting a tennis ball on a screen, it is possible to produce a good specification of the range of motion that might be expected and, using accelerometers or vision-tracking, it is possible to measure this motion from a handheld unit in order to recognize an action (which after all, is the approach taken by the Nintendo Wii). In this latter case, the person performs as action which is intended to be functionally equivalent to that performed in real-life and expects the computer to make an appropriate response, by having the avatar that the person is controlling perform the same action. Current commercial approaches to capturing such actions rely on accelerometers to respond to movements of the device held in the hand. This provides a reliable means of capturing gross movement but struggles with finer movements that might characterize many dexterous actions. Thus, in order for the capture of human movement to develop there is a need to allow computers to respond to fine motor control.

For this paper, the focus will be on ways in which one can capture the actions that people perform with objects in the real world, and treat these in much the same way as gestures are treated, i.e., as actions that can be recognized and evaluated. One way in which this can be achieved is through further refinement of the objects that the person holds when interacting with the computer. In this paper, the focus lies on capturing data from the handles of domestic tools and using these data to model different types of performance. It is proposed that such developments not only provide an interesting ground for exploring ways of analyzing human activity but also lead to the development of novel forms of interaction device. There are a number of domains in which such recognition could prove valuable, and in this paper considers two of these:

(i.) Capturing and recording everyday actions of people undergoing healthcare or rehabilitation, not in the laboratory but in their own home;
(ii.) Monitoring the actions of technicians involved in maintenance work and producing a computer log of the work.

These domains of application are considered further in the discussion section. In the next section, a review of approaches to analysing human activity from sensor data will be presented. This is followed by a discussion of capturing data from human interaction with tool handles, and a description of a prototype system. Then the results of initial trials and analysis of different activities is presented, before the paper concludes with a discussion of future developments.

1.1 Using Sensors to Analyse Human Activity

Previous research has looked at the automated analysis of ambulatory motion, with some real-time feedback, to aid in rehabilitation of walking [19][28], and at the use of activity recognition to monitor arm movement [15][17]. Such systems enable rehabilitation to be carried out at home, with the use of unobtrusive sensor systems at a reasonable cost compared to current hospital medical systems [6]. Amft and Tröster [1][2] used a range of sensors on the person to define movements involved in eating, as part of a diet monitoring application. For example, a microphone and electromyography sensor was used to recognize chewing and swallowing, and accelerometers on the lower arm indicated movements towards the mouth.
In maintenance work, Ogris et al. [30] combined a body-worn ultrasonic unit to track hand location with an accelerometer to track hand movements when people performed bicycle repair tasks. By capturing the action performed, the system was able to provide guidance and feedback to the user regarding appropriate actions to perform. Another paper reports a system in which combined RFID and bar-code reading is used to identify tools and components, and accelerometers on wrists, to define actions, with a head-mounted web-camera to record and check maintenance activities [31]. In this case, recognition was used to both guide feedback to the user and also to capture novel approaches to a task (which could then be filmed, using the head-mounted camera, for inclusion in future training videos). In a similar manner, Maurtua et al. [25] developed a system to recognize picking up a tool or component and using this to determine whether a car assembly task was being performed correctly. Stiefmeier et al. [34] defined car assembly as a series of sub-tasks and sought to recognize when each sub-task had been completed. In these papers, recognition had the primary goal of checking maintenance procedures against ‘good practice’.

### 1.2 Using Sensors to Record Grasp

In the field of ergonomics, grasp is often evaluated through grip dynamometry. This involves the person pulling against a sprung handle; the amount of force used to pull the handle is measured off a calibrated scale. This shows effective grip strength but does not provide an indication of how well the person can grasp an object or how grasp varies with activity.

The instrumentation of tools to measure grip force has been explored previously in many specific applications such as golf grip [22] and children’s handwriting [9]. The approach in these studies was to cover the handle of the tool in a force sensing mat. Both studies used the Tekscan 9811 sensor, which consists of a 0.1 mm array of force sensing cells that respond to force with a linear change in resistance. However, there are more traditional forms of sensor that are much cheaper and which could provide usable data, in the form of strain gauges. Murphy et al. [29] use strain gauges on the top and sides of a knife blade, near the handle, in order to measure forces applied during cutting. Memberg and Crago [27] designed a two sided handle, with strain gauges on each side. This design was used as the basis for the initial prototype in this paper (see figure 1). McGorry [26] used a three sided handle, with strain gauges on each side, and this was used as the basis for the design of the second prototype (figure 2).

These previous studies, regardless of the sensors used, concentrated on the design of the handle and the collection of data from the use of the sensors. However, there was little attempt at using these data to interpret the activity beyond simple visual analysis. If these devices are to be useful in HCI, there is a requirement to develop techniques for classifying and recognizing activity from instrumented tools. To this end, Kranz et al. [23] fitted a torque sensor between the handle and blade of a large chef’s knife. The data collected from this sensor, combined with the data from load-cells under a cutting board, could be used to characterise the cutting of different foods. This shows how the use of instrumented tools can provide data to support activity recognition. However, the Kranz et al. [23] study was concerned with the forces applied through the tool’s blade rather than the interaction between hand and handle. It is, therefore, of interest to ask whether hand-handle interactions can be captured with sufficient reliability to allow actions to be classified. In this paper, our aim is to model the hand-handle interactions (through motion and grip) and it would be interesting to consider whether this approach could be comparable to that used by Kranz et al. [23]. Consequently, the testing procedure that is employed requires users to perform activities using an instrumented knife; the activities include cutting different foods and spreading butter.

### 2. CLASSIFYING ACTIONS

Modelling of human performance, on the basis of accelerometer data, has been performed by neural network analysis [24][39], through hidden Markov Modelling, [3][20][42] or through Gaussian Mixture Models [31]. Each approach has the potential to be computationally intensive and in this project, the objective was to use a technique which could
run on a low power processor, so would be less computationally demanding. This could involve the use of classifiers, such as Naive Bayes and C4.5 [5] [33][36]. The features in the Naive Bayes classifier were modelled using a Gaussian distribution. Training data are used to calculate parameters that define a probability distribution for each feature in each class. These parameters form the classification model. Classification involves using a probability distribution function with the parameters of the model to calculate the probability of each feature of the unknown sample data. Naturally varying phenomena, such as human actions, tend to vary with a Gaussian distribution; hence it is deemed useful in human activity recognition. The C4.5 algorithm generates decision trees, where the leaves are the classifications, and the rest of the nodes above them are the features. The decision tree is then used to classify unknown data samples by traversing down the tree based on the value of each feature of the unknown sample. When a leaf of the tree is reached, the classification is found. In contrast to Naive Bayes, decision trees strongly model interdependence of features. The C4.5 algorithm is a well developed algorithm for building trees which deals with issues such as over-fitting data. The implementation of the C4.5 algorithm decision tree and naive Bayes classifiers is relatively simple, compared to the implementation of many of other classifiers.

2.1 Software Development

The project required a number of software modules to be developed to capture of data from the instrumented handles and perform classification. While there are various commercial packages that can do some of these tasks, it was felt that developing modules in-house would allow greater control over the manner in which the data were processed. All modules were written in C#, running under Windows.Net.

The first module was an application for real-time visualisation and recording incoming data from the sensors (Figure 3). It displayed graphs for analysing the real-time output of the three accelerometer axes, and the strain gauge output. Following the capture of data, the next step required segmentation to be done with relatively minimal effort from the user while maintaining a high level of precision (Figure 4). A second module allowed the recorded data to be visualised on a scrollable graph, and segmented and categorised simply by clicking on the graph. This allowed rapid removal of all irrelevant and null data.

A third module was written to calculate large sets of features from the samples. The features included 128 and 256 point discrete Fourier transforms (DFT), mean, variance and energy. The DFT was particularly useful due to its ability to characterise repetitive data and has been used with great success in many activity recognition applications [5] [31]. The application outputs the feature data in a format readable by the WEKA workbench.

Before classification, two feature subset selection (FSS) methods were used to remove features calculated from samples that had low salience to the classification of the data. This is important for reducing the dimensionality of the data and improving the efficiency of the classification process.
because features that do not help characterisation can dramatically reduce the accuracy of recognition. There are various approaches to FSS that can be used. Although it is theoretically possible to test every possible subset against the classification algorithm, in practice this is impractical. This project uses over 100 features resulting in over $10^{100}$ calls to the classifier algorithm (which would take years of processing time on conventional computers). For this reason, FSS methods generally use some kind of search method, which involves gradually building up the feature set using heuristics to reduce the search space. The wrapper method [21] is one of the more powerful methods and involves the use of the classifier algorithm to help evaluate the best subset. The wrapper method usually gives superior results to filter methods (methods that do not use the classification algorithm) due to the fact that they produce results specifically suitable for the classification method [16].

Filter methods use similar search methods to the wrapper method, but they do not use the classification algorithm; instead they use a function that evaluates the merit of the features against the training data. These methods tend to be much faster than the wrapper method [16] and they also have the potential to be useful with many different classification methods. Correlation-based Feature Selection (CFS) is demonstrated by [16] and shown to give significant improvement when used with a Naive Bayes Classifier (this kind of classifier is also used in this work).

Both filter and wrapper methods can have different search methods applied to them. Kohavi and John [21] show that the Best First search method, which uses some simple heuristics, generally finds better subsets than a simple greedy search. Both wrapper and CFS filter FSS were individually tested in this project using best first searches.

3. ACTIVITIES FOR CLASSIFICATION

The instrumented knife was used to perform a range of simple tasks that could be commonly used in domestic settings. These tasks include basic action on a variety of materials. The actions, and materials, were as follows:

i. cut_cheese_block
ii. cut_cheese_flat (slice)
iii. cut_cucumber
iv. cut_cucumber_flat (slice)
v. cut_orange (through peel)
vi. cut_toast vii. get_butter
viii. spread_butter_toast

Each action was performed 15 times by each of the five volunteers who participated in the data collection phase. After processing, this gave some 600 samples of data for testing.

The classifier algorithms were evaluated using hold out testing; the dataset was split into three equal sets, and then every two set combination was used as the training set, with the third used for testing. This makes sure that the test data has never been seen by the classifier; this is important because the point of a classification algorithm is to recognise unknown data. Although testing on the training data would show that the classifier is working, it would not tell you if it has the ability to cope with real data with random variation. It is possible for a classifier to get perfect classification results on the training set but completely fail in a real example because problems such as over-fitting of data would not be shown in testing against the training data.

The test set-up was the same for all participants and activities; the participant, sitting at a table, was presented with items on a plate. These items (i.e., cheese, orange, cucumber, toast) were cut or otherwise acted upon using the instrumented knife. Each activity started and ended with picking up and putting down the knife, with multiple cutting/spreading/slicing actions performed in between. All of the data that did not contain any action related data (irrelevant leading and trailing data and long pauses) were removed, splitting some of the actions into multiple samples. All the samples were automatically segmented into uniformly sized sub-samples suitable for feature creation. Each dataset used in the leave-one-out testing was acquired from separate occasions of data collection.

As Figure 5 illustrates, the accelerometer showed some variation in terms of broad type of activity. In comparison with the ‘spreading’ activity, the cutting activities generally returned very small degrees of motion as demonstrated in the second part of the above diagram.
In Figure 6, the activity of cutting a piece of toast is recorded. The uppermost plot shows variation in grip, as measured by the strain gauge, and the other three plots show movement in the three axes of the accelerometer. It can be seen that the cutting action is preceded by an increase in grip force, which is maintained until the cut has been made, and then the force reduces.

Combinations of the data from the different sensors were used to classify the actions. For example, Figure 7 clearly shows the separation of multiple activity classes by two accelerometer features; some of the classification was successfully carried out using only accelerometer features.

3.1 Recognition Accuracy

It was pointed out, in section 2.1, that the classifiers used in this study had been selected because of their relatively low computational overhead. This means that they might be expected to perform less well than more sophisticated methods. In terms of recognition performance, both the naïve Bayes and the C4.5 classifiers achieved precision and recalls above 60% which indicates that these classifiers work. On average, the naïve Bayes classifier performed better than the C4.5 decision tree classifier. Across both datasets the errors were most common between the two cheese cutting and two cucumber cutting activities, indicating that these activities are similar. If these pairs of similar classes were regarded as the same, the naïve Bayes classifier achieved precision and recall values of 90% and above.

None of the feature reduction methods improved the results when using the C4.5 classifier. The wrapper FSS method was the most beneficial, but due to its extreme computational complexity, it could take prohibitively long times to run when used on larger datasets. The CFS method only slightly improved precision and reduced recall, but the reduction of the features is still useful as it reduces the computation time. The most useful feature subsets found did not reject the features calculated from grip force, which shows that there was value to including the force sensor. However, some features calculated from the accelerometer were ranked as more valuable, indicating that a combination is required.

4. DISCUSSION

This paper demonstrates the development of a prototype instrumented handle. The use of the off-the-shelf sensors means that the device is potentially cheap to produce and the simple classifiers that have been implemented show that it is possible to determine which actions are being performed. While the recognition rates for individual actions vary from 60% to 90%+, it is likely that the handles would be used in conjunction with other sensors, e.g., RFID, which would provide additional data to support
the classification of activity. If the electronics were reduced further (e.g., through the implementation of a MEMS solution) it would be possible to embed the sensors, processor and communications entirely in the handle of the tool. The challenge lies less in the implementation of the sensing components and more in the capture and processing of the data that are produced. In terms of HCI, the concept underlying this design is to provide a means of allowing people to use familiar, everyday tools and objects in their normal environments. This allows them to focus on the physical tasks that they would normally perform, with a computer being able to record specific actions. In a previous study of maintenance work [31], we demonstrated how the capture of activity concerning user movement, from sensors on the person, and RFID could be used to both generate sets of instructions for performing the tasks (in the form of training videos for uncommon tasks) and the logging of actions (which could be compared against a job-list or procedures). This paper shows how it might be possible to have the sensors fitted on the tools that a person uses, which we argue would be less intrusive than having the sensors on the person. In terms of rehabilitation, the ability to capture behaviours in the person’s normal and familiar environment, in terms of re-learning simple domestic tasks, could prove an interesting and beneficial development. Having a means of capturing sensor data and classifying specific actions could provide an indication of changes in performance. In terms of maintenance, the ability to monitor tool use could not only provide a way of tracking performance (and comparing this against the standard procedures that need to be followed, particularly in safety critical systems) but also to assess condition and wear of tools or level of ability of the tool user. This information could form part of tool-replacement program in preventative maintenance or an indication of the need for refresher training of personnel. It may be possible to recognise anticipatory grip force before the user starts different phases of the activity. In familiar situations, where an increase in load is predictable, e.g., when picking up an object, grip force is typically adjusted in phase with changes in load [12][13][41]. Studies such as [18] and [37] show that grip force adjustments in holding a tool, prior to a collision, anticipate the impact force in terms of velocity. These studies imply that people adjust their grip force, on the handles that they are holding, in anticipation of future actions or effects. This notion could be used to further refine the modelling processes, e.g., either in terms of structuring the activity into phases, or in terms of defining the sequences with which actions are performed. We could then compare the time spent in anticipation or action across different types of user or different conditions. This could, for example, provide a fine-grain measure to compare performance over time in order to see if the performance of the user has improved, perhaps as the result of practice or training.

The focus of this paper has been on the use of simple classification schemes to label particular tasks performed with instrumented handles. This can provide a record when tasks were performed (by logging them in a time-stamped database), perhaps for monitoring maintenance work or for recording everyday behavior in a home-setting for rehabilitation. Further work can be applied beyond the simple classification to consider the performance of individual tasks. For example, by comparing the pattern of activity performed by an individual against a template representing ‘good’ performance, it is possible to compare experts against novices (in maintenance work) or to evaluate changes in performance (in rehabilitation). While these analyses are beyond the scope of this paper, the prototypes and data collection capabilities we have developed will support this as the next stage of development for the work.

5. REFERENCES

[6] Bonato, P., 2005, Advances in wearable technology and applications in physical medicine and rehabilitation, Journal of Neuroengineering and Rehabilitation, 2,
and rehabilitation 87, 1542-7
[18] Johansson, R.S. and Westling, G., 1984, Roles of glabrous skin receptors and sensorimotor memory in automatic control of precision grip when lifting rougher or more slippery objects, Experimental Brain Research, 56, 550-564
[38] Underkoffler, J. and Ishii, H., 1999, Urp: a luminous- tangible workbench for urban planning and design, CHI '99, New York: ACM, 386-393

© 2010, the Authors