FoodBoard: Surface Contact Imaging for Food Recognition

Peer-reviewed author version

Pham, Cuong; Jackson, Daniel; SCHOENING, Johannes; Bartindale, Tom; Plötz, Thomas & Olivier, Patrick (2013) FoodBoard: Surface Contact Imaging for Food Recognition. In: Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing, p. 749-752.

DOI: 10.1145/2493432.2493522
Handle: http://hdl.handle.net/1942/15823

FoodBoard: Surface Contact Imaging for Food Recognition

Cuong Pham[1], Daniel Jackson[2], Johannes Schöning[3], Tom Bartindale[2]
Thomas Plötz[2], Patrick Olivier[2]

[1]Post & Telecom
Institute of Technology
Hanoi, Vietnam
pcuongcntt@gmail.com

[2]Culture Lab
School of Computing Science
Newcastle University, UK
{firstname.lastname}@newcastle.ac.uk

[3]Hasselt University
tU-iMinds
Diepenbeck, Belgium
johannes.schoening@uhasselt.be

ABSTRACT
We describe *FoodBoard*, an instrumented chopping board that uses optical fibers and embedded camera imaging to identify unpackaged ingredients during food preparation on its surface. By embedding the sensing directly, and robustly, in the surface of a chopping board we also demonstrate how surface contact optical sensing can be used to realize the portability and privacy required of technology used in a setting such as a domestic kitchen. FoodBoard was subjected to a close to real-world evaluation in which 12 users prepared actual meals. FoodBoard compared favourably with existing unpackaged food recognition systems, classifying a larger number of distinct food ingredients (12 incl. meat, fruit, vegetables) with an average accuracy of 82.8%.

Author Keywords
UbiComp, Sensing Surfaces, Food Recognition.

General Terms
Human Factors; Design; Measurement.

INTRODUCTION & MOTIVATION
Recognizing food in everyday contexts, such as kitchens, is a canonical problem for the ubiquitous computing research community [15], and an important component of systems that could potentially undertake automated dietary intake estimation, provide situated support for healthier eating habits, or automated assistance for meal preparation. Previous research in food recognition has typically used computer vision techniques, i.e. physically remote camera imaging [10,11,13,17]. However, given the intrinsic complexity of the problem, even approaches that significantly restrict the context (by constraining the image capture configuration and/or range of foods) yield relatively unsatisfactory recognition rates. In one of the most successful examples, Yang *et al.* [17] utilised a fast food database and used pair-wise local features to represent the spatial relationship between pixels of different food types. Yet even in this highly constrained context with only seven distinct food types (e.g. *sandwich*, *salad & sides*, *bagel*, *donut* and similar) SVM-based classification yielded only a 78% accuracy rate.

As an alternative to computer vision based approaches, and partly in response to inevitable concerns about users' privacy, a small number of food recognition systems have been based on acoustic data. For example, Amft *et al.* [1] used an in-ear microphone to classify different types of food based on the chewing sounds. Using a decision tree classification for 4 texturally distinct food items (*chips*, *apple*, *pasta*, and *lettuce*) a two-step acoustic analysis gave an accuracy of between 66-86% overall for a "single chew", and between 80-100% for a "chewing segment". However, identification of food as it is about to be swallowed is too late for many applications. By contrast Kranz *et al.* [12] developed a multimodal food recognition system that combined a microphone installed in their Aware Kitchen, and a knife augmented with force and torque sensors. An overall recognition rate of 85% accuracy for 6 food items demonstrated the potential for their food classification, although the obtrusive nature of the technologies used clearly indicated the need for more sensitive product design if such technology-enhanced utensils are to be used in real-world settings.

For classifying food in everyday settings such as a kitchen, traditional configurations of computer vision and acoustic sensing fall short on two related counts: *(i)* they involve obtrusive sensing technology (i.e., cameras or microphones); and *(ii)* they pose significant privacy concerns for the people that use the space in which they are deployed. While RFID readers embedded in kitchen worktops have been shown to address these privacy concerns, they are unable to identify foods which do not have packaging in which tags can be embedded; in particular, fresh food ingredients such as meats, vegetables and fruits. In response to this challenge we describe *FoodBoard*, a pervasive sensing and context recognition system developed to recognize unpackaged ingredients, as they are prepared on a chopping board. *FoodBoard*'s novel surface contact imaging technique is embedded within a custom-designed chopping board, thereby alleviating any potential privacy concerns and allowing the board to be freely moved and washed.

Our design goal for FoodBoard was to develop a modular component of a monitoring system for domestic activities

UbiComp'13, September 8–12, 2013, Zurich, Switzerland.

http://dx.doi.org/10.1145/2493432.2493522

(in this case for aspects of food preparation).Food recognition results are intended to inform a higher-level inference system that provides, for example, situated prompting to support the autonomy of people with cognitive impairment [6]), task-based language learning [7] or development of cooking skills [16]. In these application scenarios the embedded nature of FoodBoard's sensing has inherent advantages over more generically configured computer vision systems (e.g., an overhead camera) as: *(i)* FoodBoard only senses objects and activities on the surface of the board and thus more fully respects the privacy of people in the kitchen; and *(ii)* by not relying on modifications to the infrastructure of the kitchen (as camera setups required for occlusion free capturing would) its deployment requires little or no more effort than that of a regular chopping board.

FoodBoard

(i) Design and Construction

There is a technical challenge in creating an embedded sensor capable of recognizing food: a certain minimum sensor area and resolution is required. Even with a special lens for a camera to image the underside of the chopping surface a greater distance between the camera and the surface is required than the thickness of a chopping board allows. While custom electronics incorporating many hundreds of (costly) colour optical sensing elements could be used, we instead utilize the approach used in FiberBoard, the first compact surface imaging sensor board based on fiber optics [8]. In FoodBoard a matrix of optical fibers (**Error! Reference ource not found.**) is used to channel light from the sensing surface to a camera attached below. Channelling light with optical fibers permits flexible placement of the camera to maintain a reasonable overall thickness for the chopping board (in practice this is limited only by the minimum bend radius of the fiber).

Unlike FiberBoard, for which infrared light is channelled from a multi-touch surface, FoodBoard channels visible light from a support surface beneath the chopping board's transparent upper surface. The measurement grid contains 600 holes in a 30×20 matrix covering an area of 315×210 mm, which is representative of the active area of chopping boards in everyday use. An optical fiber is inserted into each of the holes and the unattached ends of all fibers are gathered together into a bundle (50×30 mm). A camera is arranged so that it directly faces the bundle to capture the light reflected along the fibers by food on the chopping board surface. Synthesizing the image for recognition requires a mapping to be constructed (during calibration) from the image of the fiber bundle to the sensor grid. This arrangement produces a compact on-surface imaging device, which intrinsically promotes privacy as a result of the extremely local imaging range and resolution.

(ii) Calibration of Mapping and Image Synthesis

The optical fiber bundle is incoherent, in that it has no defined mapping between the fiber bundle and the sensing grid. Nevertheless, such a mapping must be produced in order to synthesize the images sensed by the receiving end of the fibers. This is achieved through a calibration procedure. An effective calibration and synthesis mechanism significantly enhances the utility of low-cost fiber bundles, and allows the development of a whole class of thin form factor image processing applications..We modified the calibration routine described in [9] to fit to our prototype. First, a visually opaque rectangular tool (e.g. a ruler) is moved at a roughly constant velocity in a direction perpendicular to each of the two orthogonal axes of the board's surface. At the same time the camera's images of the optical fiber bundle are recorded. The timing, in both passes, of the resulting fluctuating light from each fiber now approximately identifies the sensing coordinates of each image pixel.

Next, pixels from a single fiber tip must be grouped together – yet this cannot easily be done in the plane of the camera image (as fiber tips with similar mapping results can be immediately adjacent, making them difficult to distinguish). Instead, the destination mapping is reverse-projected from the timing values, "blob detection" is used to label connected regions (the sensing grid), and the regions grown up to neighboring regions. Finally, these regions are used to group the pixels of a single fiber.

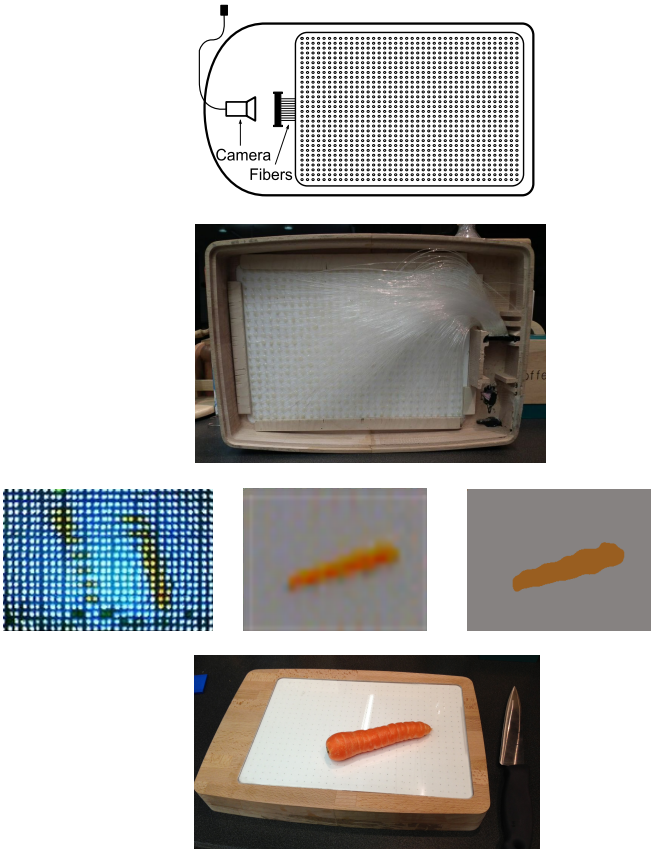

Figure 1. *FoodBoard* schematic (top); Underside interior view of bundled fibers during construction (second row); Example captured image of the fiber bundle, reconstructed image using the calibrated mapping, and segmented result (third row, left-to-right); *FoodBoard* in use (bottom)

At run-time, the input mapping is applied to the camera image, giving the average colour and brightness for each fiber. The individual fiber point values are then interpolated to synthesize the sensed image, which is then used as the input to the final processing stages: segmentation, smoothing, feature extraction and matching. See Figure 1 for initial, intermediate and final images of the fiber bundle.

(iii) Optical Food Recognition

The synthesized image of the chopping board surface is first processed in two steps to optimize the food recognition: segmentation and smoothing. As the number of colours of a natural food image is unknown, an unsupervised method is necessary to perform food image segmentation. Therefore, K-Means Clustering algorithm was implemented to perform real-time colour segmentation. When captured under realistic settings the images contain significant noise, which gives rise to some (small) clusters unconnected to the region of interest. To resolve this we applied a morphological closing-opening operation to smooth the segmented image. The main region of interest that contains food will be extracted using colour and SURF features for food recognition (in the next step).

The final step to be performed is the actual food recognition step based on the processed image generated by the FoodBoard fiber-based surface contact imaging system. The recognition algorithm utilizes SURF and colour features to classify food images and was implemented to allow for real-time food recognition.

Feature extraction

Speeded Up Robust Features (SURF) [2] is widely known as one of the most robust feature detectors and is used in numerous object tracking [5,14] and object recognition [4] applications. SURF is also well known to handle serious blurring. Moreover, SURF features are also invariant to rotation and scale. These characteristics are very important for classification of food ingredients processed on the chopping board, as the position of food will vary and they will have different sizes (e.g. a "baby" carrot is essentially a smaller version of a "large" carrot). While a SURF descriptor includes the discriminating features one might expect, such as angle, edges and points, the standard SURF ignores colour. However, colour information is also very important for discriminating between food ingredients so, in addition to SURF features, we also extract colour features and use these in our recognition pipeline.

To classify food, a feature extractor (FE) was implemented that comprised two main procedures. One is an implementation of Fast-Hessian detector, and the other is an RGB colour histogram. The input of FE is an image segmented by our implementation of the K-Means Clustering algorithm, the output of which is two lists: one contains a 64-element list of SURF interest points (SURF features) $S=(s_1, s_2, ..., s_{64})$, and the other is a 64-element color histogram $C=(c_1, c_2, \ldots, c_{64})$. After normalization, these lists are combined into a 128-element feature vector:

$$V = [\alpha * s_1, \alpha * s_2, .., \alpha * s_{64}, (1-\alpha) * c_{65}, (1-\alpha) * c_{66}, \ldots, (1-\alpha) * c_{128}]$$

where α is a weight by which SURF and colour matching features are proportionately ranked. In our experiment, a value of α=0.4 was heuristically chosen (by evaluating different values of α in a pilot study). Thus colour features are slightly more heavily weighted than SURF features.

Matching

Two algorithms were compared for food image classification: k-Nearest neighbour (k-NN) and Support Vector Machines (SVM). The former was implemented from scratch for real-time food recognition and the latter was based on libSVM [3]. The reasons for choosing k-NN and SVM are that both these algorithms can deal well with high dimensional data (i.e. 128-d feature vectors) and both afford real-time implementation. Furthermore, SVM is able to deal effectively with an unbalanced dataset (a typical characteristic of real-world datasets).

As irrelevant objects unintentionally placed on the chopping board (i.e. a knife or a user's hand) must be rejected, a threshold t_i was assigned for each training food class f_i. The algorithm parameters (i.e. threshold) were manually defined as the result of a 4-fold cross-validation procedure on the Euclidean distance between test and training food images. Such a food recognition algorithm is simple, but fast enough for real-time image classification. A food is matched if the distance of at least one of k closest images is greater than a threshold for the food class; otherwise, f_i is rejected (i.e. classified as an unknown food).

EVALUATION

A study was conducted to evaluate the food recognition algorithms performance of FoodBoard under realistic conditions. We asked 12 participants to cook a full meal with all ingredients we provided. The preparation phase involved washing and chopping the ingredients on the board. The collected dataset is comprised of 1,800 images of 12 food ingredients. The number of each type of food image was randomly selected between 50 and 250. The collected food images were naturalistic, in that they varied in object position, rotation, included hand "distractions", and notably many also contained water-base reflections. The evaluation results from this study are presented in Table 1.

Results

As can be seen in Table 1, *bacon* and *carrot* have high recognition rates (over 90%). While the colour of bacon and carrot are similar, their SURF features are quite distinct, and consequently only very few of the images of carrots and bacon were misclassified. Images yielding large numbers of false positives (i.e. over 20%) included *peeled onions* and *yellow peppers*, thus the misclassification of (and confusion between) *peeled onions* and *yellow peppers* was significant, as was the case for *tomatoes* and *red peppers.* In a few instances, sticks of celery and leaves of *lettuce* were misclassified as a result of having very similar colour and SURF features. However, the overall recognition precision

Ingredient	Precision	Recall
Bacon	85.23 %	96.15 %
Carrot	90.08 %	92.18 %
Celery	59.09 %	87.64 %
Cucumber	87.30 %	88.00 %
Dill	88.27 %	78.61 %
Green pepper	85.29 %	88.46 %
Lettuce	93.20 %	67.71 %
Onion	75.00 %	60.78 %
Red pepper	73.30 %	81.64 %
Spring onion	83.33 %	87.12 %
Tomato	89.20 %	69.27 %
Yellow pepper	73.50 %	72.08 %
Mean	**82.76 %**	**78.77 %**

Table 1: *FoodBoard* evaluation results.

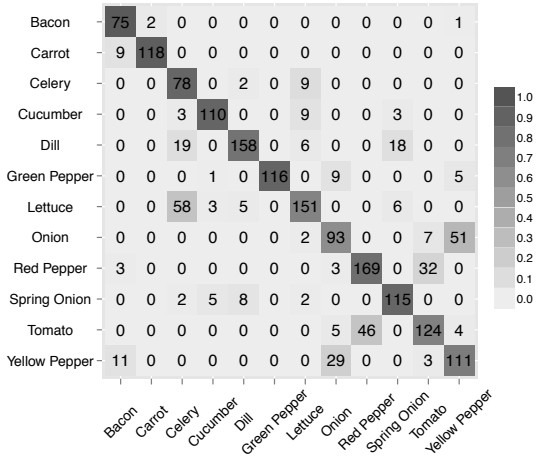

Figure 2: Confusion matrix for food recognition evaluation

and recall rate was approximately 80% on 1,800 images of 12 food ingredients thereby demonstrating that the fiber-based surface contact imaging method has genuine promise as a practical embedded food recognition technology.

CONCLUSION & DISCUSSION

We have presented FoodBoard, an augmented chopping board that uses fiber-based surface contact imaging technology to automatically recognize unpackaged food ingredients. We have also demonstrated the viability of FoodBoard as a potential building block of kitchen activity monitoring systems and its advantages over traditional direct sensing approaches in terms of the privacy conserving nature of the embedded imaging, the modular nature of the chopping board (i.e. not requiring modifications to the kitchen environment) and more robust recognition.

Our first FoodBoard prototype is comparable to a regular chopping board in terms of size, weight, and functionality, but the camera integrated into the *FoodBoard* for our evaluation study was wired (with a USB connector). This shortcoming has subsequently been addressed through the use of an integrated wireless webcam; although this design still presupposes a central computer capable of performing the requisite imaging processing operations we have described. What recognition performance is required for a genuinely useful unpackaged ingredient recognition technology such as *FoodBoard* can only be determined by reference to both the nature of the situated support or monitoring required by the overarching application in which it is a component, and by understanding the level of variety of unpackaged ingredients that users of *FoodBoard* might actually use. However, we have presented the design and development of a novel augmented appliance with embedded context-recognition, including our justification for the selection of materials and hardware, its design and construction, image calibration, processing and food recognition algorithms.

REFERENCES

1. O. Amft, M. Stäger, P. Lukowicz, and G. Tröster. (2005). Analysis of chewing sounds for dietary monitoring. In *Proc. UbiComp.*
2. H. Bay, A. Ess, T. Tuytelaars, and L. van Gool. (2008). Speeded-Up Robust Features (SURF*). Comp. Vis. Image Underst.* 110(3), 346-59.
3. C. Chang and J. C. Lin. (2011). LIBSVM : a library for support vector machines. *ACM Trans. Intell. Systems and Technology*, 2(27) 1-2.
4. J. Chen, A. H. Kam, J. Zhang, N. Liu, and L. Shue. (2005). Bathroom activity monitoring based on sound. *In Proc. Pervasive.*
5. R. Chincha and Y. Tian. (2011). Finding objects for blind people on SURF features. In *Proc. BIBMW*.
6. J. Hoey, T. Plötz, D. Jackson, A. Monk, C. Pham, and P. Olivier,. (2011). Rapid specification and automated generation of prompting systems to assist people with dementia *Pervasive and Mobile Computin*g (PMC), 7(3), 299-318.
7. C. Hooper, A. Preston, M. Balaam, P. Seedhouse, D. Jackson, C. Pham, C. Ladha, K. Ladha, T. Plötz, and P. Olivier. (2012). The French Kitchen: Task-Based Learning in an Instrumented Kitchen. In *Proc. UbiComp*.
8. D. Jackson, T. Bartindale, and Patrick Olivier. (2009). FiberBoard: compact multi-touch display using channeled light. In *Proc. ACM ITS.*
9. T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu. (2002). An Efficient k-Means Clustering Algorithm: Analysis and Implementation. *IEEE TPAMI,* 24(7), 881-892.
10. K. Kitamura, T. Yamasaki, and K. Aizawa. (2009). FoodLog: capture, analysis and retrieval of personal food images via web. In *Proc. ACM Multimedia Workshop Multimedia for Cooking and Eating Activities*.
11. L. Kok-Meng., L. Q. Qiang, and D. Wayne. (2007). Effects of Classification Methods on Color-Based Feature Detection With Food Processing Applications. *IEEE T-ASE,* 4(1), 40-51.
12. M. Kranz, A. Schmidt, B.Rusu, A. Maldonado, M. Beetz, B. Hornler, and G. Rigoll, (2007). Sensing Technologies and the Player-Middleware for Context-Awareness in Kitchen Environments. In *Proc. INSS.*
13. G. Shroff, A. Smailagic, and D. P. Siewiorek. (2008). Wearable context-aware food recognition for calorie monitoring. In *Proc. ISWC.*
14. D. N. Ta, W. C. Chen, N. Gelfand, and K. Pulli. (2009). SURFTrac: Efficient Tracking and Continuous Object recognition using Local Feature Descriptors. In *Proc. CVPR.*
15. P. Olivier, G. Xu, A. Monk, J. Hoey. (2009). Ambient kitchen: designing situated services using a high fidelity prototyping environment. In *Proc. PETRA.*
16. J. Wagner, A. van Halteren, J. Hoonhout, T. Plötz, C. Pham, P. Moynihan, D. Jackson, C. Ladha, K. Ladha, and P. Olivier. (2011). Towards a Pervasive Kitchen Infrastructure for Measuring Cooking Competence. In *Proc. PervasiveHealth*.
17. S. Yang, M. Chen, D. Pomerleau, and R. Sukthankar. (2010). Food recognition using statistics of pairwise local features. In *Proc. CVPR*