Material change for
a better environment

Final Report

Using Machine Vision to Sort Plastics

A pilot study in detecting milk bottles during the HDPE recycling process

Project code: ISS001-001
Research date: Sept 2010-Mar 2011

Date: September 2011

WRAP helps individuals, businesses and local authorities to reduce waste and recycle more, making better use of resources and helping to tackle climate change.

Written by: Ioannis Alexiou and Dr Anil Anthony Bharath, Department of Bioengineering, Imperial College London

Front cover photography: Plastics recycling at MRF

WRAP and Imperial College believe the content of this report to be correct as at the date of writing. However, factors such as prices, levels of recycled content and regulatory requirements are subject to change and users of the report should check with their suppliers to confirm the current situation. In addition, care should be taken in using any of the cost information provided as it is based upon numerous project-specific assumptions (such as scale, location, tender context, etc.).
The report does not claim to be exhaustive, nor does it claim to cover all relevant products and specifications available on the market. While steps have been taken to ensure accuracy, WRAP cannot accept responsibility or be held liable to any person for any loss or damage arising out of or in connection with this information being inaccurate, incomplete or misleading. It is the responsibility of the potential user of a material or product to consult with the supplier or manufacturer and ascertain whether a particular product will satisfy their specific requirements. The listing or featuring of a particular product or company does not constitute an endorsement by WRAP and WRAP cannot guarantee the performance of individual products or materials.

Executive summary

Background:

Plastic milk bottles are a good source of high-density polyethylene (HDPE) that occurs in large quantities in household recycling material. Existing systems for sorting plastics rely on near-infra red (NIR) sorting machines, which can sort based on polymer type (e.g. HDPE or polyethylene terephthalate (PET)). Some plastic containers and bottles in the recycle stream might have similar NIR signatures to milk bottles but could have been used to hold contaminating materials such as household cleaners. These containers can appear identical to the milk bottles under NIR examination. Human sorters are currently required to perform separation of food-grade from non food-grade HDPE – this human element currently presents a bottleneck to the capacity of rHDPE plants.

Objectives & Approach:

This report presents the findings of a 6-month project to investigate the feasibility of using machine vision techniques to identify food-grade HDPE, exemplified by plastic milk bottles, during the recycling process. The objectives of the project were to i) acquire data and to assess the practicality of using the visual appearance of items on the recycling conveyor to identify the food-grade HDPE items ii) to suggest an appropriate real-time computer vision architecture for detecting the milk bottles iii) to estimate the achievable detection accuracy in terms of false positives and false negatives in such a detection system iv) to provide a non-real-time demonstrator of a training and detection system. This report summarises the techniques used to explore the feasibility of automated item classification in a recycling setting and presents initial findings.

Findings:

The high degree of deformation of the milk-bottles during the recycling process means that a 100% rate of detection is unlikely. Although preliminary, experimental work suggests that a system for achieving good sorting with very low false acceptance in labelling food-grade items could be achieved; such a system would need to incorporate an extensive and updateable training process, whereby three primary visual characteristics are used to detect and sort the bottles under controlled lighting and acquisition conditions. These are: i) visible-light colour signatures ii) shape signatures with a complexity related to the number of items used in the training process and iii) visible labels validated against an up-to-date product label database. Whilst i) and ii) increase the yield of detected milk bottles, the use of labels to validate and detect objects would carry two strong advantages: (a) it would incur a shorter training process, should new products appear into the recycling chain and (b) there is a much lower chance of detractor items, collapsed into similar shapes to those of compacted food-grade items, being misclassified.

Further Work:

Further work which could prove valuable and interesting in this area following this trial includes the:

- Marking up of more video data to both train and assess accuracy;
- Development of modules/algorithms to incorporate flexible (non-rigid) spatial grouping of features;
- Development of a training process to allow non-expert users to train an assembly-line system;
- Real-time harvesting of low-level visual descriptors.

Contents

Figures

Tables

Acknowledgements

The authors would like to thank WRAP for funding the project. Special thanks to Lynne McGregor and Brian Graves (Imperial Innovations) for setting up and supporting the work. Thanks to Brian Castelino of Machine Vision Technology for discussions and work on data acquisition. Nick Cliffe (Closed Loop Recycling) for provided invaluable advice, data and support during the running of this project. Thanks to Lucas Hadjilucas (Imperial College London) for his support on the use of SVMs.

1.0 Background

New processes now enable plastic bottles to be recycled to produce food-grade packaging. Nearly 20% of the plastic bottles currently produced in the UK can now be recycled, reducing the quantities of both landfill and exported material. The process is currently economically viable, but margins, yields and quality must be carefully monitored.

During the recycling process, it is important that plastics that are either not food-grade, or that have been in contact with potential contaminants, are appropriately binned. For example, coloured HDPE bottles are separated and sold to manufacturers of drainage pipes, wheelie bins, garden furniture and other items. Colourless HDPE often, but not always, is suitable for food-grade recycling. Many of the plastics made from HDPE and suitable for food-grade recycling are sourced from milk bottles, and tend to be largely colourless. Manual sorting plays a critical role in the food-grade plastics recycling process, particularly in identifying the items that are appropriate for food-grade recycling.

It is generally considered that deformable objects, such as plastics that undergo high degrees of physical warping encountered in a plastics recycling plant, present a challenge to image recognition systems. This limitation has, for example, made the creation of systems that can interpret human, or other "biological" sources of visual motion, difficult without the use of tracking equipment, visible markers or biomechanical models of visual motion.

Is it possible to develop a system for plastics sorting at speed that is capable of working at least as well as a human sorter? The fact that it is relatively easy for humans to detect certain types of household plastic receptacle on the recycling conveyor belt is an indication that the task should be achievable using existing technology; human visual perception is essentially a computational process: neurons in the human visual system respond to visual stimuli in early development, and plasticity in these neurons alters the spatial patterns of neuronal connections and/or their connection strengths, such that their firing patterns provide a coding of visual inputs. This constitutes a training process, during which synaptic connections are strengthened or pruned to allow the visual system to learn to **represent** the visual information seen by the eyes, and thereby **discriminate** between visual patterns.

Accordingly, an approach that might be taken to develop a system to recognise and, ultimately, to sort food-grade containers should incorporate both components: technology to provide appropriate descriptions of visual patterns of recycling waste and a learning approach to discriminate the visual patterns associated with significant items.

2.0 Approach

In order to assess the feasibility of detecting plastics in the process of recycling from their visual appearance, the following steps were taken:

- Four visits to the Closed-Loop Recycling plant in Dagenham were undertaken to observe and record the flow of items during the recycling process of a live, running, plastics recycling plant.

- A separate investigation into the feasibility of using shape, colour and label detection in a lab-based setting was also undertaken in order to assess the possibility of employing a training process based on item shape, colour and label appearance.

- An assessment of the necessary camera resolution, imaging speed and lighting to acquire suitably discriminating images of items during the recycling process was undertaken in a live recycling plant.

- Purchase, construction, and temporary installation of a prototype image acquisition system. This included installation of controlled lighting and a customised video camera rig capable of fast shuttering to capture video sequences of items during the recycling process in order to minimise motion blur in an active plant.

- An assessment of the relative frequency of occurrence of a known, high-yield, HDPE plastic occurring in milk bottles in captured sequences was estimated by sampling frames at intervals over a 1 hour time frame.

- A manual process to mark up the locations of milk bottles in training and test sequences was performed; this was used to estimate the accuracy of bottle detection using different subsets of features of visual appearance.

- Initial work has been undertaken to develop a prototype training system that synthesizes rotation and minor scale changes in images of objects to increase the efficiency of the training process in learning important invariant properties.

- Prototype algorithms have been developed for extracting colour and shape features from items on the conveyor belt during recycling; these have been customised to the visual recognition problem posed by the plastics recycling context.

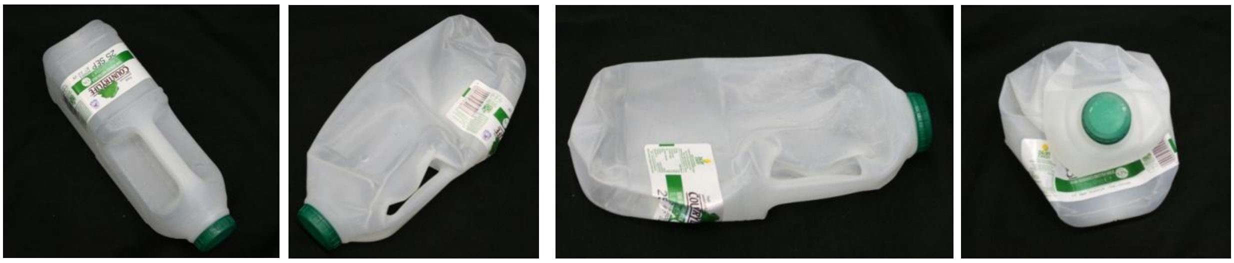
3.0 Preliminary Data Acquisition

Following initial studies in a laboratory setting using different types of source cameras, a recommendation was made on the resolution required for data acquisition in the recycling plant. This recommendation enabled a bootstrapping process to be performed: gathering data enabled an evaluation of the potential accuracy of bottle detection in a commercial setting. It also yielded further recommendations on the data acquisition process for acceptable sorting levels.

3.1 Laboratory setting

A number of sample images of HDPE milk bottles were obtained, and roughly but loosely crushed to assess the stable characteristics of the bottles. Using a digital SLR camera, a number of photographs of the bottles were taken to assess variability under typical conditions encountered in the plant. As shown by **Figure 1**, although the appearance of the milk bottle is clearly identifiable, the shape characteristics, as assessed by the bottle outline, can be quite variable. For example, aspect ratios can vary significantly.

Figure 1 Sample images of HDPE milk bottles. The range of deformations is typical of the distortions encountered in a recycling setting; nevertheless the items are recognisable as milk bottles partly by their translucence, with supporting cues of colour (particularly the caps) and validated by distinctive labels and/or features.



The use of high quality SLR images of target items such as those shown in **Figure 1** are suitable for theoretical tests of absolute detection capability of objects in a laboratory setting, and for building high-quality appearance-based models (Donner, 2006) ; they do not provide a good indication of real-world performance of detection algorithms.

For this reason, a second source of laboratory images was acquired using a standard, but low-quality Web camera (Logitech) under variable lighting conditions. Using these images, an initial assessment was made of the required resolution to adequately detect distinctive visual features on the bottles. For example, using a standard SIFT-based keypoint approach (Lowe, 2004) the minimum spatial resolution needed to allow discrimination to be performed was determined to be no less than 1.6pixels/mm on the surface of the object to be classified. Furthermore, difficulty in stabilising distinctive points on the surface of the object to be classified was identified as a key problem to be addressed in an industrial setting. More details are available in **Section 4**.

3.2 Recycling Plant

Based on the assessment of image quality and availability of keypoints in the laboratory setting, a prototype test rig was constructed and installed over two days at the Closed Loop Recycling Ltd, (Dagenham, Essex, RM9 6LF). The rig consisted of 4 high frequency lights and a CCD camera of 1280x960 image resolution, aligned so that the long axis spanned the wide axis of the conveyor belt (900mm). The camera was repositioned for some image sequences to increase the image resolution up to around 1.7pixels/mm. A number of image sequences were collected, amounting to over 120,000 frames of image data. Different frame rates, object belt distances and belt speeds were used; the trade-off in detection performance for different configurations, belt speeds and acquisition rates have potential impacts on the cost of acquisition hardware needed to obtain acceptable performance. These data are comprehensive but represent >3GB of data, and are yet to be comprehensively analysed, primarily because significant manual mark-up of the images is required in order to assess accuracy. A table describing the available statistics on acquired data is given in **Appendix A**. A sample of a typical section of the data containing over 300 frames of image data is also presented in **Appendix B** and testing data presented in **Appendices C** and **D**.

3.3 Recommendations based on preliminary data acquisition

The following requirements are suggested as a *minimum* to enable recognition to be done:

- Controlled lighting – required to minimise lighting variations in the field of view of cameras and throughout the day for plants that are partially exposed to daylight variations. This lighting should be diffuse to avoid specular reflections, and arranged to minimise shadows.

- Fast shuttering on the video camera to minimise object blur; 1/10,000 s is suggested.

- A spatial resolution for image acquisition **that is at least 2.5 pixels per millimetre on the surface of the object to be categorised**. Lower resolutions are permissible (1pixel/mm) when the object is to be discriminated by gross shape. However, to protect against inaccurate detection, increased spatial resolution is advisable to allow texture signatures to be captured on the surface of the object. This is discussed further in **Section 4.**

- A rate of acquisition that ensures that each object on the conveyor belt is captured *fully* at least ONCE in the frame, and preferably TWICE. An increased number of acquisitions could improve detection rates in terms of false positives and false negatives.

4.0 System Requirements: Computer Vision

4.1 Introduction

Because it is critical that the system should err on the side of rejecting plastic receptacles that have not contained food – even if the plastic is food-grade HDPE – label or object part recognition should be used as part of the process. Identifying parts of known labels is, perhaps, the most reliable visual indicator of whether the receptacle contained food or not: if the label is recognisable as a known food product, it is more likely to be safe for food-grade material recycling from the point of view of freedom from contamination.

Whilst it would be feasible to achieve some recognition success based on shape and colour, the severe deformations experienced in recycling and the lack of distinctiveness of colours for food grade material **suggests that label recognition, based on patch descriptors, for example, should be employed to "gate" final decisions.** Thus, even if an item were to pass basic tests, such as having a plausible colour signature, and a size and silhouette typical of a 2L milk bottle, accepting it as usable for food-grade recyclable material should require that the label also be recognised. The downside to this is that items containing severely damaged or missing labels will be rejected.

4.2 Literature Review of Relevant Techniques

There are numerous techniques that could contribute to the detection of milk bottles during the recycling process. However, key requirements of the system should be scalability, the ability to rapidly retrain, and the potential to minimise computations for speed of operation without a high degree of hardware customisation. Overviews of techniques that are widely used in the industrial vision context are provided by Davies (Davies, 2005), and Sonka (Sonka, 2007). More advanced approaches to problems of scalable recognition are covered in papers on object recognition in large databases (Nister, 2006) and video retrieval (Zisserman, 2009). Earlier work on geometric hashing (Wolfson, 1997) also suggests fast mechanisms to learn geometric relationships between identifiable parts, particularly where one wishes to validate a hypothesis that has been put forward by a retrieval algorithm earlier in the recognition process. The use of colour is covered in numerous basic texts, but descriptors that use joint colour and spatial properties would be powerful; see, for example, Schiele (Schiele, 1996).

4.3 A Recommended System Design

Initial experiments described in food-grade plastic receptacles could be identified by the following sequence of operations:

- Locating individual items on the conveyor belt using background subtraction or intensity-based thresholding. This may fail in black plastics, but we have not encountered high incidence of these in the test sequences so far analysed.

- Calculating colour signatures for each item. Eliminating certain objects based on known colour profiles.

- Partitioning items to locate distinctive parts, such as colourful caps/covers that help identify the object – e.g. milk bottle caps.

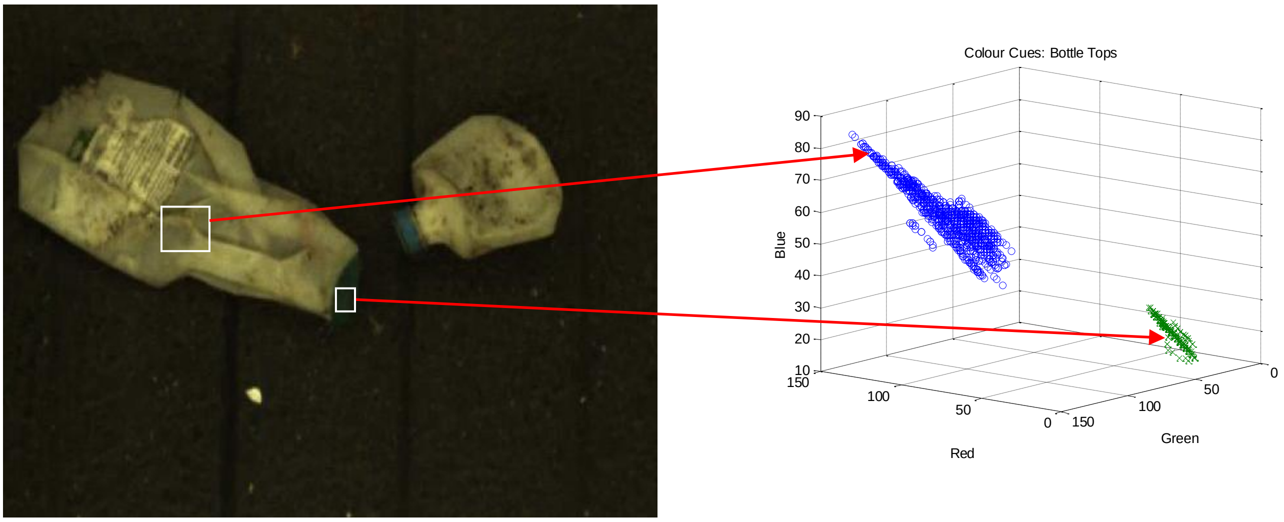
- Use of simple shape descriptors to eliminate implausible objects: examples would include aspect ratios that are unlikely to be milk bottles.

- Use of label reading, and an up-to-date product label database to provide high accuracy verification for individual items.

4.4 Use of Colour

Although colour plays an important role in the sorting process within plastics recycling, the use of colour is a first step in distinguishing plausible recyclable items. For example, although the items shown in **Figure 2** are not easily discernible, the item on the left has a high probability of being a milk bottle because of the presence of a green cap (later removed in the recycling process). Examples of two items on the conveyor belt of a recycling plant are shown in **Figure 2**.

Figure 2 An example frame showing items on the conveyor belt during plant operation. High-speed shuttering is used, and lighting is controlled. Extreme deformation of the objects is visible. The labelled regions illustrate the colour information available. There is sufficient separation in colour signatures to contribute towards item recognition, though colour alone, without spatial properties, is not a sufficient visual cue.

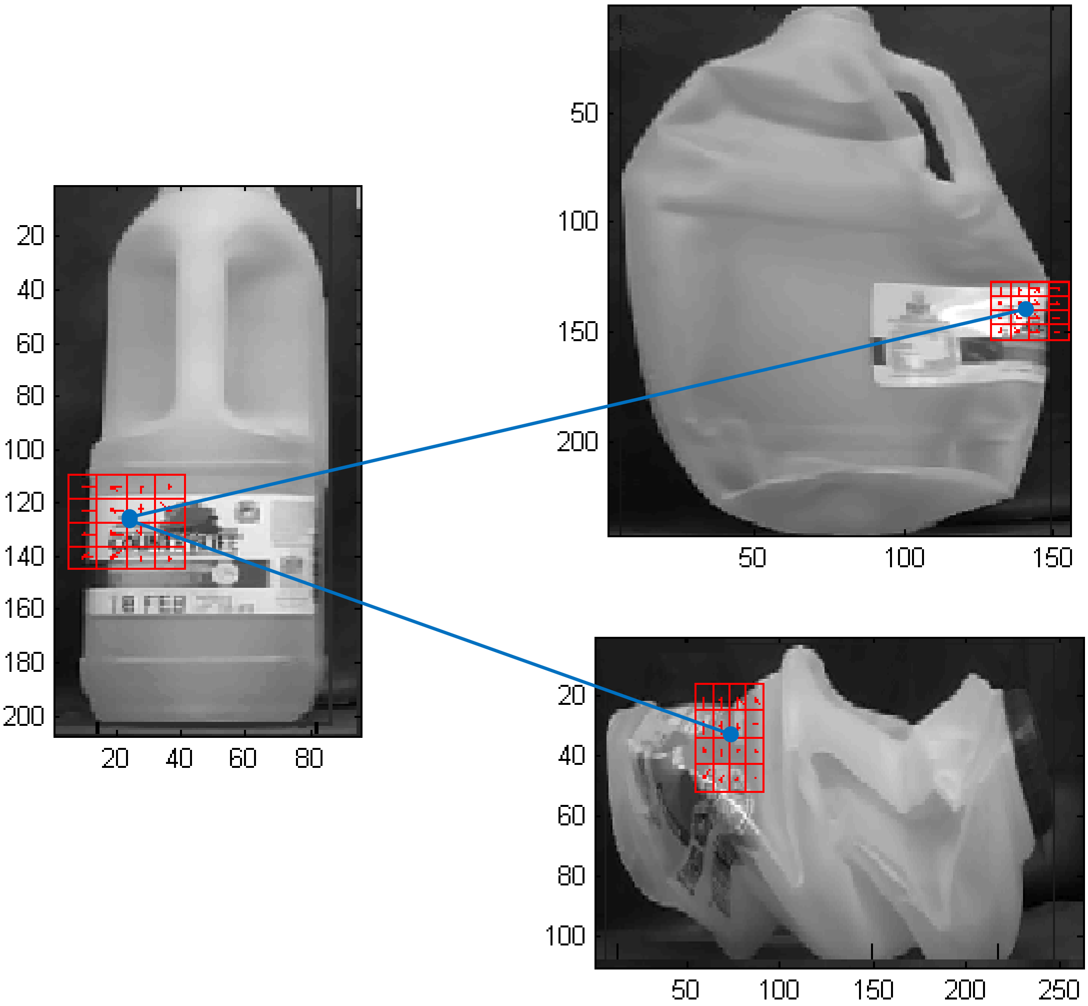


Experiments on the use of colour have demonstrated that colour signatures alone can achieve around a 70% accuracy rate in classifying milk bottles. This could be increased if colour descriptors for an entire candidate object are combined with colour descriptors for individual patches, as suggested in **Figure 2**. Sufficient training data and part relationship modelling is required to achieve this. **Table B** provides an indication of the achievable rates using colour histograms alone.

4.5 Matching distinctive points

At a low level, the patches, or regions that have been identified on the surfaces of objects that are known to be of interest – target items – can be sought on query items that are localised in the field of view of the camera. This principle is illustrated in **Figure 3**.

Figure 3 Patches on labels of known items can often be found on query items, even in the presence of distortions, provided that the local distortion, over the scale of centimetres, say, is not too great. The human visual system is very good at identifying these distinctive, stable points, and so it is appropriate to incorporate such matches to boost the level of confidence in accepting, or rejecting, an item. Many industry standard algorithms describe patches by their local structure.



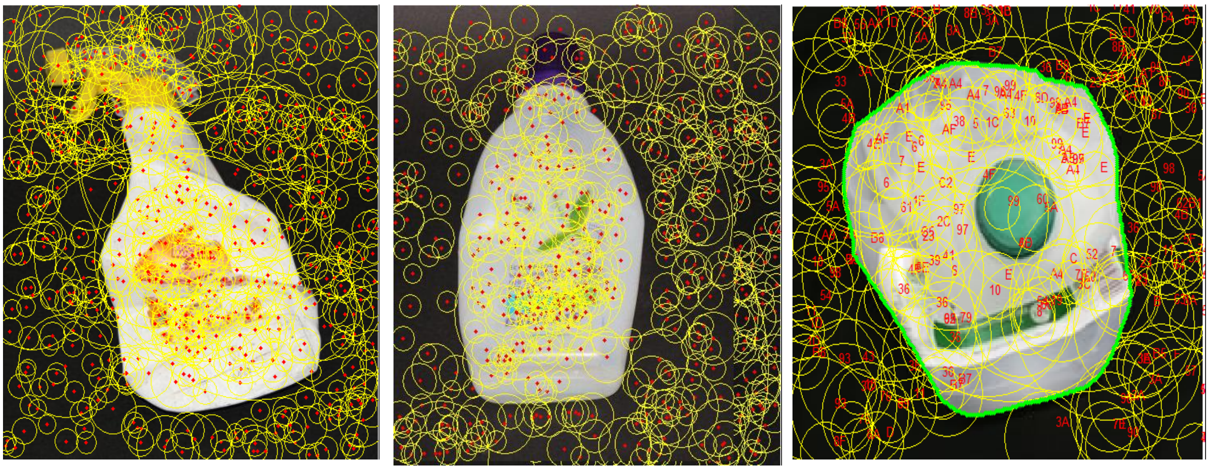
4.6 Shape-Based Approach Using Keypoints

The keypoint approach, as outlined in (Lowe, 2004), provides a way of identifying small or large coherent visual cues, including boundary cues, that are distinctive of a target object. **Figure 4** illustrates such focus-of-attention points in a lab-based acquisition and training system. Note the presence of points in the background as well, which are removed in subsequent processing based on background subtraction – only the points that are interior to the shape boundary are used for training or discrimination.

The idea of focus-of-attention contributes towards the ability to perform object recognition in real-time against a large database of previously seen examples by reducing the number of points that is extensively described. Although recognition accuracy can usually be improved by the use of more keypoints, the scalability that is afforded means that there is a certain amount of flexibility available: increasing the numbers of keypoints may require more hardware, but yields better recognition performance.

Keypoint methods are generally accompanied by descriptors – collections of numbers that characterise a local region, often in a manner that is tolerant to scale or rotation changes.

Figure 4 Illustration of focal attention points (red dots) of an industry standard keypoint detector (SIFT), and approximate scales of distinctiveness (yellow circles). The rightmost figure shows a further two stages of processing. In the first, the main (coarse-scale) boundary is found for the image, then cluster IDs are assigned to each keypoint descriptor. Fine scale textures that are associated with labels, portions of objects (such as handles) and distinctive large scale curvatures that can be considered as visual cues.



To achieve scale-invariant performance, regions within the circles of **Figure 4** are all re-scaled to the same canonical size, then described using a 128 element vector (Lowe, 2004). These 128 element vectors are then *clustered* (Nister, 2006) to form *visual words*, and histograms of visual words can then be used to detect conjunctions of image patches that are highly indicative of a target object being present in the field of view.

The key locations, scales and keypoint descriptors are used to build up a vocabulary of visual structure which can be used for subsequent searches. In order to do this, only the visual words – quantised descriptors – that are associated with valid content are learned. For example, a first step in teaching a system to recognise particular shapes of interest would involve keeping only the visual words that are located within a certain region of the image, such as those falling within a bounding polygon of a region identified in frame.

4.7 Interior/Exterior Points

In examples shown on the rightmost panel of **Figure 4**, the words that are **within** the green boundary (automatically determined) would be kept, and assigned to the class "detractor", whilst those outside the boundary are assigned to the class "background" (i.e. not an object of interest at all). The boundaries of shapes, obtained from training data from samples of objects to be classified, for example, can typically be determined automatically using standard computer vision approaches and controlled image acquisition; this simplifies the training process. Only visual structures that are guaranteed to occur *consistently* for either "detractor", "background" or "target" would ever be preserved. Others would be discarded as being irrelevant to the sorting task.

5.0 Estimated Achievable Performance Rates

5.1 General Remarks

A certain percentage of distinctive label features are absent on recyclable bottles, due to occlusion, being completely face down such that the compressed sides of the bottle (where labels often wrap around) are also obscured. This accounts for approximately 20% of bottles, though it is difficult to be certain without more extensive and time-consuming manual image labelling arranged by sampling more extensively the trial sequences. It is recommended that such bottles, which cannot be verified by known labels, should be classified as non-milk bottles, or "unverifiable".

It is feasible that other signatures of milk bottles that are based on colour and shape characteristics might be attainable using machine learning (Zhang, 2006) and employing on the order of tens of thousands of target examples. However, such an approach will only be as good as the training data is representative of the samples encountered during normal running of the recycling system.

Labels, on the other hand, need far fewer examples because of their generally distinctive appearance; retraining in the instance of changes to the products that appear in the recycling chain is then a much less onerous process, and can even be fed from information about product labels available from third parties.

For example, BrandBank (http://www.brandbank.com/) maintains databases of images of product labels which also contain linked product data. This product information may be linked to product category information that should be indicative of possible contaminants. For example, common bottles containing household cleaning products are usually registered on BrandBank; because the images on BrandBank are used for supermarket web sites and internet shopping sites, the images held by BrandBank are linked to product Ids.

5.2 Experimental Results: Summary

The experimental results for training, and testing performance using colour only are presented in **Tables B** and **C** in the Appendices. Manually labelled ground truth images were used to provide training data and to assess the accuracy on independent test data. For shape and label descriptions, the visual word vocabularies were built using a single pass clustering applied to standard SIFT descriptors. It is generally known that larger numbers of descriptors, and richer vocabularies obtained from more extensive training data generally lead to increased performance (Nister, 2006). No geometric validation of spatial relationships between keypoints was used; incorporating this will further increase the accuracy of classification.

Based on the preliminary experiments with an early system incorporating SIFT descriptors and colour histograms, but without colour parts indexing or any form of geometric verification, and using in nearly 400 frames of test and training data, covering more than 800 objects that were manually labelled, the **accuracy** obtained, defined by,

$$Accuracy\ (\%) = \frac{True\ Positives + True\ Negatives}{True\ Positives + True\ Negatives + False\ Positives + False\ Negatives} \times 100$$

is **74%-76%.**

This is a conservative estimate using the available training data, and does not take into account other measures that might be more important in a recycling context. For example, it is possible to set an operating point to decisions that guarantees that only items that are certain to be milk bottles should be so labelled. This is best assessed by considering the measure known as **False Discovery Rate (FDR):** a measurement that indicates the fraction of objects classified as milk bottles which actually are not milk bottles. Keeping the FDR as low as possible should be the goal of the classification algorithm in order to minimise contamination of the food-grade HDPE.

Our estimates of the FDR at different stages of the system are as yet preliminary. Better gauging of accuracy rates cannot be attempted before extensive manual markup is performed, but the results are certainly promising, with an FDR in testing data at or below **0.2**, even with the limited quantity of system development and training undertaken to date.

It is very likely that with more extensive training – for example, by doubling the number of training examples – and increasing the size of the visual word vocabulary (Nister, 2006) which would require markup of training data, that accuracy rates **in excess of 85%** would be achievable. It is **estimated**, based on knowledge of visual word vocabulary size and known performance figures with increased learning (Nister, 2006), that an FDR of less than **0.1** should be achievable within 6 months, and lower than **0.05** achievable within 12 months, once geometric verification, via hashing or spatial pyramids (Lazebnik, 2006), is incorporated.

6.0 Feasibility of Developing a Commercial System

It is likely that a commercially available machine vision platform could be modified to include the necessary keypoint features to allow the label-based recognition to be added to standard colour and gross shape-based metrics. A major component of the design should be the inclusion of a customised training module, possibly fed from label and object databases, but including the ability for plant operators to re-train they machine every one or two months with minimal effort. Building this is likely to constitute a main part in the cost of development.
A commercial system would at least need to incorporate modules as shown in the flowchart in **Figure 5.**

Figure 5 A proposed flowchart for processing the video sequence. Although many of these components can be found in off-the-shelf machine vision systems, on-line supervised training is an essential component in the recycling context; most components in the chart above will have to be optimised due to the nature of the recycling context, in which the items being imaged are highly variable – unlike standard industrial vision settings in which traditional vision algorithms operate.

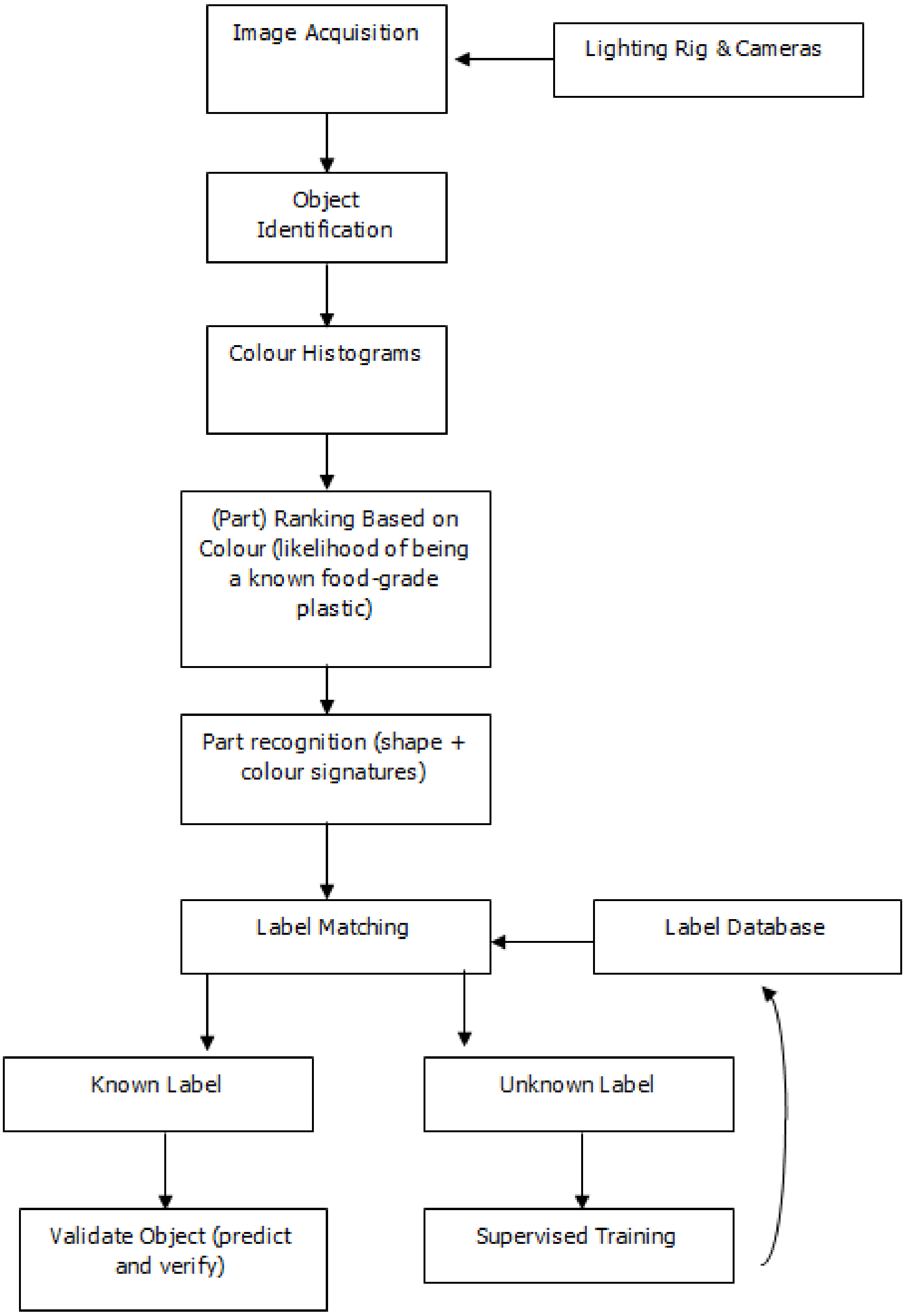


Image Acquisition
Lighting Rig & Cameras
Object Identification
Colour Histograms
(Part) Ranking Based on Colour (likelihood of being a known food-grade plastic)
Part recognition (shape + colour signatures)
Label Matching
Label Database
Known Label
Unknown Label
Validate Object (predict and verify)
Supervised Training

Although the demands of recognising deformable objects have required expensive configurations in the past, the use of the streamed processing capabilities of GPGPU (General Purpose Graphical Processing Units) means that hardware costs are now dominated by camera and lighting costs. However, GPGPU systems carry a premium in programming costs, so that a cost/benefit trade-off would need to be undertaken.

7.0 Conclusions

Although preliminary for a vision task of this complexity, the work to date does suggest that it will be possible to build a system using image recognition to bring down costs, and to increase throughput and yield in food-grade plastics recycling. Such a system would require extensive training on manually labelled data, and should be built so that the bias in detection is on ensuring the positive identification of packaging from known food-containing products. Labels of "negative" or inappropriate products could also be used to reject packaging that could contain known contaminants. In both cases, the positive identification of a product is highly desirable.

Custom computer vision techniques generally incorporate most of the components needed for a real-time system. It is likely that some custom components would need to be developed in engineering a system that achieves the best performance. A key requirement, however, is that any system would need to be easily retrained in order to accommodate modifications in product packaging, and the appearance of new objects in the recycling chain.

Even the best sorting system based on visual recognition could still carry risks of plastic contamination. For example, food containers that are re-used by householders to store cleaning products, detergents etc, can still present cases where contamination is possible, even if it is likely to be rare. A system for improving the recycling throughput and reducing dependency on manual sorting will still require regular assessment of the food-grade quality of the plastics produced by the recycling process.

8.0 References

Davies, 2005. ER Davies, *Machine Vision : Theory, Algorithms, Practicalities*. Morgan Kaufmann, ISBN 0-12-206093-8.

Donner, 2006. R Donner, M Reiter, G Langs, P Peloschek and H Bischof, 2006; Fast Active Appearance Model Search Using Canonical Correlation Analysis, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.28, no.10, pp1690 - 1694.

Lazebnik, 2006. S Lazebnik, C Schmid and J Ponce; Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2006*, Vol.2, pp2169 - 2178.

Lowe, 2004. DG Lowe; Distinctive Image Features from Scale-Invariant Keypoints, *International Journal of Computer Vision*, Vol. 60, pp91-110.

Nister, 2006. D Nister and H Stewenius; Scalable Recognition with a Vocabulary Tree, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2006*, Vol.2, no., pp 2161- 2168.

Philbin, 2010. J Philbin, M Isard, J Sivic and A Zisserman; Descriptor Learning for Efficient Retrieval**,** *Proceedings of the European Conference on Computer Vision*.

Schiele, 1996. B Schiele and J Crowley; Object recognition using multidimensional receptive field histograms *ECCV '96, Lecture Notes in Computer Science*, Springer Berlin / Heidelberg, Vol. 1064, pp610-619.

Sivic, 2009. J Sivic and A Zisserman; Efficient Visual Search of Videos Cast as Text Retrieval, *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Volume: 31 Issue: 4, April. pp591 - 606.

Sonka, 2007. M Sonka , V Hlavac and R Boyle. *Image Processing, Analysis and Machine Vision*, 2007.

Wolfson, 1997. HJ Wolfson and I Rigoutsos; Geometric hashing: an overview, *Computational Science & Engineering*, IEEE , Vol.4, no.4, pp10-21.

Zhang, 2006. Hao Zhang, AC Berg, M Maire and J Malik; SVM-KNN: Discriminative Nearest Neighbor Classification for Visual Category Recognition, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2006*, Vol.2, pp2126 – 2136.

Appendix 1: Summary of Data Set 1

This appendix summarises one of two data sets. At time of writing, the second data set has not yet been received.

Table 1 A summary of the data collected on Day 1 of the data acquisition. "Far view" and "close view" refer to two different distances of acquisition, corresponding to resolutions of approximately 1.2pixels/mm and 1.6pixels/mm, respectively.

	Number Of Files	Total Duration	Frames	Frame Rate	Total Size	
Directory 1 Far View	21	01:35:47	45976	8	2,33 GB	LEAD codec
	1	00:01:59	1428	12	119 MB	Using LEAD codec
	1	00:10:04	5436	9	18,5 GB	Uncompressed
	1	00:00:23	299	13	1,05 GB	Uncompressed
Directory 2 Close View	1	00:00:15	180	12	180 MB	Uncompressed
	1	00:02:27	2058	14	7,11 GB	Uncompressed
	1	00:00:09	162	18	629 MB	Uncompressed
	1	00:00:12	228	19	822 MB	Uncompressed
	4	00:08:53	11726	22	40,2 GB	Uncompressed

Appendix 2: Training Data — Colour

This appendix summarises statistics of images over nearly 300 image frames. It also presents **training** accuracy once a representative colour histogram has been learned for milk bottles. Overall accuracy is **72%**. FDR: **0.18.**

Table 2 Image training data over 1 second intervals of video at 8fps. Ground truth was manually determined by a single human observer, and therefore also subject to error. Last column shows training classification rate achievable. The 5 rightmost columns containing classification numbers using colour histogram distances to classify the training data. T_0 and T_1 are times in seconds defining frame intervals.

		# of	# Milk	# Non Milk	True	False	False	True	
T_0	T_1	Objects	Objects	Objects	Positives	Positives	Negatives	Negatives	Accuracy
0	1	24	18	6	17	1	1	5	0.92
1	2	16	13	3	9	0	4	3	0.75
2	3	13	13	0	13	0	0	0	1.00
3	4	13	6	7	4	4	2	3	0.54
4	5	13	8	5	4	2	4	3	0.54
5	6	8	7	1	5	0	2	1	0.75
6	7	14	9	5	9	4	0	1	0.71
7	8	14	10	4	8	3	2	1	0.64
8	9	10	8	2	7	2	1	0	0.70
9	10	9	6	3	5	3	1	0	0.56
10	11	8	5	3	4	2	1	1	0.63
11	12	14	10	4	8	2	0	3	0.79
12	13	12	9	3	8	3	1	0	0.67
13	14	13	7	6	6	4	1	2	0.62
14	15	11	9	2	8	0	1	2	0.91
15	16	10	9	1	8	0	1	1	0.90
16	17	10	6	4	4	2	2	2	0.60
17	18	13	10	3	7	1	3	2	0.69
18	19	15	11	4	9	2	2	2	0.73
19	20	20	14	6	9	3	5	3	0.60
20	21	6	5	1	4	0	1	1	0.83
21	22	5	4	1	4	0	0	1	1.00
22	23	12	10	2	7	0	3	2	0.75
23	24	9	8	1	5	0	3	1	0.67
24	25	11	9	2	7	0	2	2	0.82
25	26	6	5	1	4	1	1	0	0.67
26	27	14	10	4	8	1	2	3	0.79
27	28	13	11	2	7	0	4	2	0.69
28	29	10	8	2	6	0	2	2	0.80
29	30	15	12	3	11	1	1	2	0.87
30	31	15	10	5	6	3	4	2	0.53
31	32	17	13	4	10	2	3	2	0.71
32	33	13	9	4	7	2	2	2	0.69
33	34	12	9	3	8	1	1	2	0.83
34	35	8	6	2	4	2	2	0	0.50
35	36	11	7	4	7	1	0	3	0.91
36	37	11	6	5	4	5	2	0	0.36
TOTALS		448*	330	118	261	57	67	62	

* One object is missed due to occlusion.

Appendix 3: Testing Data — Colour

This Appendix summarises **testing** accuracy once a representative colour histogram has been learned from the training data presented in **Table 2**. Overall accuracy is **71%**. FDR: **0.23.** Accuracy of sorting is high, but FDR is higher than for training data.

Table 3 Image testing data over 1 second intervals of video at 8fps. Ground truth is manually determined by a single human observer, and therefore also subject to error. The 5 rightmost columns contain classification numbers using colour histogram distances to classify the testing data. The increase in the FDR relative to the training data is not desirable; in particular, one wishes as low a FDR as possible, because it is an indicator of non-milk bottles (detractors) that have been classified as milk bottles. T_0 and T_1 are times in seconds defining frame intervals.

T_0	T_1	# of Objects	# Milk Objects	# Non Milk Objects	True Positives	False Positives	False Negatives	True Negatives	Accuracy
0	1	20	11	9	8	6	3	3	0.550
1	2	12	8	4	6	3	1	1	0.583
2	3	14	13	1	12	1	1	0	0.857
3	4	10	9	1	6	2	2	0	0.600
4	5	19	13	6	12	4	1	2	0.737
5	6	14	12	2	7	1	5	1	0.571
6	7	11	4	7	4	3	0	4	0.727
7	8	14	9	5	7	0	2	5	0.857
8	9	12	9	3	8	1	1	2	0.833
9	10	13	7	6	6	3	1	3	0.692
10	11	13	7	6	6	1	1	5	0.846
TOTALS		152*	102	50	82	25	18	26	N/A

* One object is missed due to occlusion.

Appendix 4: Testing Data — Shape & Colour

This Appendix summarises **testing** accuracy once representative colour histograms and a Bag of Words has been learned from the training data presented in **Table 2**. Overall accuracy is **74%**. FDR: **0.21.** FDR is reduced by using multiple visual cues. Average accuracy over second-long intervals is **76%**.

Table 4 Image testing data over 1 second intervals of video at 8fps. Ground truth is manually determined by a single human observer, and therefore also subject to error. The 5 rightmost columns contain classification numbers use colour histogram and visual word histogram distances to classify the training data. T_0 and T_1 are times in seconds defining frame intervals.

T_0	T_1	# of Objects	# Milk Objects	# Non Milk Objects	True Positives	False Positives	False Negatives	True Negatives	Accuracy
0	1	20	11	9	8	7	3	2	0.500
1	2	12	8	4	6	2	1	2	0.667
2	3	14	13	1	12	1	1	0	0.857
3	4	10	9	1	5	2	3	0	0.500
4	5	19	13	6	12	4	1	2	0.737
5	6	14	12	2	7	1	5	1	0.571
6	7	11	4	7	3	4	1	3	0.545
7	8	14	9	5	7	0	2	5	0.857
8	9	12	9	3	8	1	1	2	0.833
9	10	13	7	6	6	3	1	3	0.692
10	11	13	7	6	6	2	1	4	0.769
TOTALS		152*	102	50	97	26	4	15	

***** One object is missed due to occlusion.

Waste & Resources
Action Programme

The Old Academy
21 Horse Fair
Banbury, Oxon
OX16 0AH

Tel: 01295 819 900
Fax: 01295 819 911
E-mail: info@wrap.org.uk

Helpline freephone
0808 100 2040

www.wrap.org.uk/plastics

recycle