Project title:	Vegbot project, Robotic touch, sense, and learning of delicate vegetables
Project number:	CP172
Project leader:	Luca Scimeca
Report:	Annual report, November 2020
Previous report:	December 2019
Key staff:	Luca Scimeca Fumiya Iida
Location of project:	University of Cambridge Engineering Department
Industry Representative:	G's Growers Ltd Barway, Ely, Cambridgeshire CB7 5TZ Jacob Kirwan
Date project commenced:	01.10.2017

DISCLAIMER

The results and conclusions in this report are based on an investigation conducted over a one-year period. The conditions under which the experiments were carried out and the results have been reported in detail and with accuracy. However, because of the biological nature of the work it must be borne in mind that different circumstances and conditions could produce different results. Therefore, care must be taken with interpretation of the results, especially if they are used as the basis for commercial product recommendations.

AUTHENTICATION

We declare that this work was done under our supervision according to the procedures described herein and that the report represents a true and accurate record of the results obtained.

[Name] Fumiya Iida

[Position] Reader in Robotics

[Organisation] University of Cambridge, Department of Engineering

Signature .... Date .....24.11.2020................

[Name]

[Position]

[Organisation]

Signature .... Date ....

Report authorised by:

[Name]

[Position]

[Organisation]

Signature .... Date ....

[Name]

[Position]

[Organisation]

Signature .... Date ....

CONTENTS

GROWER SUMMARY

Headline

Automatic detection and learning for post-harvest operations in conveyor belts.

Background

Conveyor belts are core components in horticulture, where produce must be selectively discarded or reorganized. The need of tailored systems discourages solutions based on common data collection and labelling, as well as vision systems based on known object features.

Summary

Often, workers need to operate on conveyor belts, selectively picking objects based on visual features following inspection. We devise a novel framework to learn from existing human labour, without need of explicit data gathering or labelling. The framework autonomously detects, tracks and learns the visual features of salient objects in conveyor belt-based systems. The system trains entirely through visual observation of human labour and achieves detection accuracy of over 97% on a set of 7 different objects after only 10 minutes of operation

Financial Benefits

The potential financial benefits arise from two main prospects: first, the possibility of maintaining high productivity even when manual labour work force is lacking; second, the increase in efficiency and transferability (no need to train re-train labour force) possible given the potential in high precision solutions delivered by robotics platform, and their ability to work over regular work hours.

Action Points

Partial or incremental integration of the system in current pipelines would be needed. The product, however, needs to be tested in real-world settings before doing this.

SCIENCE SECTION

Introduction

Conveyor belts are core components in horticulture, where produce must be selectively discarded or reorganized. Some workers in fruit packaging or redistribution facilities, for example, would stand by a conveyor belt, while selectively picking out produce which seems damaged or not up to market standards; others, instead, might inspect mechanical components to discard defective items.
In many such scenarios workers may be working in closed off environments, with noise level exceeding 85 dB [1], at times with acidic smells coming from vinegar based solutions, rooms with temperatures lower than -20C, to preserve produce for as long as possible [2], [3], and/or low-lighting conditions [4]. More importantly, conveyor belt system tasks are often repetitive and they are cause of many injuries, first amongst many: arm and hand injuries, cuts and scrapes, burns and abrasions and bone fractures [5]–[7]. In such cases it is desirable to find automated solutions, where robotics systems can be employed to work in human-unfriendly environments.
The automation of conveyor-belt based systems in industrial settings has persevered for over 40 years [8]–[10]. In the last two decades there have been a number of robotics and Machine Learning solutions aimed at fully automatizing the industry. Work in visual servoing has been amongst the most active research areas [11], with solutions spanning from visual tracking to robot control [12]–[14].

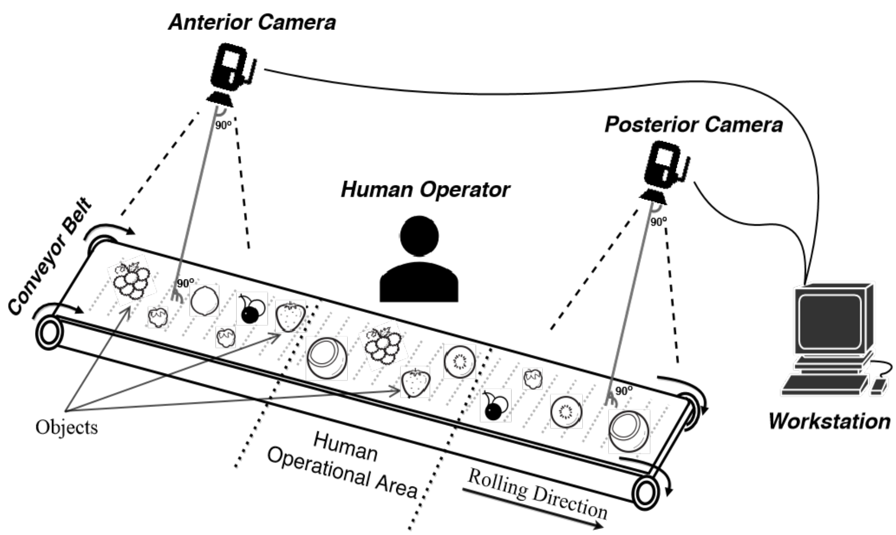


Figure 1: Scene observation and item identification framework.

The advancement in robotic gripper mechanisms has also boosted automation [15]. More recently, advancements in material science and robotics gave way to the advent of soft-robotics and soft robotics grasping mechanism [16], revolutionizing picking and manipulations tasks for a wider variety of tasks, such as picking garbage from a garbage disposal facility [17] or grasping soft fruit objects [18], [19]. The majority of the work, however, is aimed at mechanical or robotics solutions capable of detecting known objects, picking or manipulating them within the conveyor system. In several industrial scenarios, instead, the intervention on the objects needs to be selective. In a fruit factory, for example, the labour's task might be that of removing defective produce, while in a garbage disposal factory, the task might be to separate objects into different containers depending on the materials that compose them. In these scenarios it is first necessary to learn the labour operational task, to recognize and detect salient items to be acted upon.

The automation of the selective detection of items from a conveyor belt, based on task understanding, is a most useful step forward towards the full automation of conveyor belt based systems. We argue that, to be applicable in industry, a solution should have three features: first, it should need minimum data, and fast learning and retraining procedures, to limit transferability issues when re-using the same system in a different industry; second, it should ideally need no explicit data gathering and labelling, as it would otherwise be necessary to create a specific labelled data-set for each scenario, even within the same industry, limiting the portability and usability of the system by many users with ease; third, it should reach human-level selective item detection and identification accuracy, thus providing an advantageous substitute to current labour in such applications. It is also desirable for the framework to perform objects detection without any prior knowledge of object-dependent features, else a new vision-based solution must be devised for different scenarios.

We devise a novel framework to autonomously detect and track salient objects in conveyor belt-based systems. Through the use of two cameras it is possible to track elements in the belt which were acted upon by the labour, and learn without any explicit labelling or supervision to detect salient objects in future scenarios, based on the observed labour's task. Furthermore, a deep-network architecture is devised to cope with learning from streaming data, while maintaining the ability to generalize well for oncoming streams.

Materials and methods

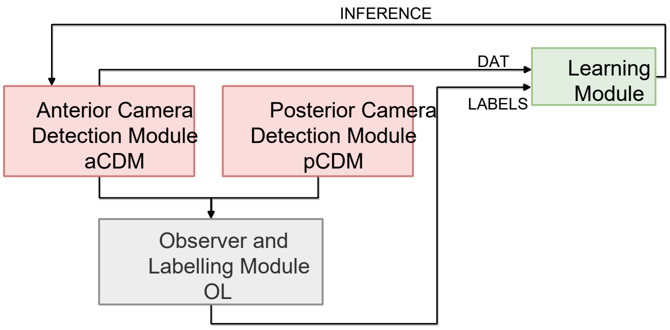


Figure 2: The Observer and Learning architecture of the developed framework

We develop a framework capable of learning to detect and select objects through visual observation of skilled labour. The system was designed to learn from minimum training data, need no explicit data gathering and labelling as conventional deep-network frameworks, and reach a human-like levels of accuracy in the detection and identification of salient objects. The framework developed is composed by a vision system and a deep learning system running on a local workstation, and is summarized in Fig 1. The vision systems is composed by two cameras and an object detection and tracking module. The system's role is that of observing and analysing human intention within the conveyor belt area. The cameras, namely the anterior and posterior cameras, are placed in two different

locations within the same conveyor belt system (Fig. 1). The anterior camera captures an area of the conveyor belt which was not affected by human intervention, whilst the posterior camera captures an area of the belt where human labour has already intervened. By comparing the analysis of the same conveyor section, prior and post human intervention, it is possible to localize the areas of the belt which were affected by human input.
Fig. 2 shows the framework's architecture. For each camera there exists a detection module, namely the Anterior Camera Module or aCDM, and the Posterior camera module or pCDM, whose role is that of localizing every object within its field of vision. The detections for each camera are sent to an Observer and Labelling Module or OLM, which compares their visual feeds to annotate those objects which were influenced by human intervention. The camera captures aCDM, together with additional information supplied by the OLM, can then be fed to a Learning Module, which trains to recognize which objects in the conveyor belt needed to be picked from at the time when they were observed in the anterior camera. All through the run, the Learning Module can make inferences on which objects need intervention and improve over time by observing human labour.

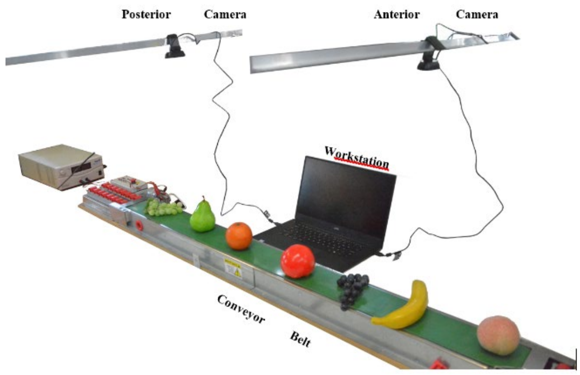
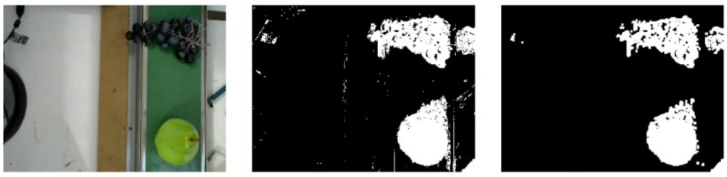


Figure 3: The Figure shows the set-up for the experiments

A. Camera Set-Up

Within the framework it is important that the two cameras' visual fields do not overlap i.e. each visual field should either be prior or post human intervention, but in no in-between state. Moreover, a stationary assumption is necessary, i.e. the position of an object within a conveyor belt is assumed not change except due to external perturbations (e.g. human intervention). In most scenarios, this can trivially be held true with small changes the belt system.
Figure 3 shows the set-up developed to validate the framework. Two low-cost Logitech C270 webcams are used, capturing visual feeds at a maximum of 30 fps, and at a resolution of 640x480 px. The two cameras are set-up 90 cm apart, facing directly downward a custom-made conveyor belt unit. The conveyor belt component has a 130 x10 cm flat upper surface, hosting a belt moving at a speed of approximately 50 cm/s.

B. Object Detection and Tracking



(a) RGB (b) Gaussian (c) Binary Mask

Figure 4: The figure shows an example detection by a CDM. (a) The sample I(i) captured frame, (b) the corresponding M(i) Gaussian probability mask, with higher intensity indicating pixels with higher probability of being part of the foreground and (c) generated

1) Camera Detection Module: Both the aCDM and the pCDM modules need to be able to detect unknown items within their field of vision. It is here desirable to devise an object

(a) RGB (b) Binary Mask (c) Boxed Detection (d) Reshaped Object 1 (e) Reshaped Object 2

Figure 5: The figure shows the image reprocessing steps before learning. From a sample I(i) frame, a (b) binary mask B(i) is created. The mask is used to (c) frame the objects, and finally (d)-(e) a number of images equal to the number of objects is created, where the objects are cropped, rotated and padded from $I^{(i)}$ to become ready for learning.

extraction technique which does not rely on any context-dependent visual features, otherwise an ad-hoc visual tracking solution must be tailored for different scenarios in each industry. First, it is necessary to make a motion assumption, i.e. we assume that if an object is placed on the conveyor belt, and is within the visual field of either the anterior or posterior camera, it must change its relative position to the cameras over time. As any one object will be in motion when the belt is operational, this will trivially be true for most systems. A foreground-extraction object detection and tracking algorithm is then devised, based on the Gaussian Mixture-based Background/Foreground Segmentation Algorithm background subtraction [20].

For each camera detection module, there is a stream of captured frames. Consider $I_a^{(i)}$ the i^{th} captured frame of the aCDM, and $I_p^{(i)}$ the i^{th} captured frame of the pCDM. Each frame is a 640x480x3 RGB array sampled at constant time intervals t_r, here $t_r = 0.03s$. The progress of i is therefore consistent with the time lapsed since the start of the system. For the sake of notation the a and p subscripts are dropped when the methods apply to both the anterior and posterior camera detection modules.

At each time interval i, a binary mask $B^{(i)}$ is generated, corresponding to the foreground of a captured $I^{(i)}$ (Fig. 4). A Gaussian mask $M^{(i)}$ can be computed as:

$$M_{(i)} = GMM(I_{(i)}) \qquad (1)$$

where GMM is the foreground extractor functions implemented in [21], based on the Gaussian Mixture-based Background/Foreground Segmentation Algorithm in [20] and [22]. Here, $M^{(i)}$ is a 640x480 array, where each element corresponds to the probability that the corresponding pixel within $I^{(i)}$ belongs to the foreground (Fig. 4a and 4b). The foreground of certainty is thus extracted, or $B^{(i)} = M^{(i)} == MAX(M^{(i)})$ where $B^{(i)}$ is a 640x480 foreground binary masks (Fig. 4c). When the motion assumption previously mentioned holds, pixels within moving objects will naturally have low probability of being part of the background, and thus will generally be present within the extracted foreground.

2) Observer and Labelling Module: The OLM module's role is to compare two frames corresponding to the same sliding conveyor belt section, prior and and post human intervention. The comparison allows for the detection of those objects within the belt which were influenced by the system.

At time i, the OLM module can observe any captured frame $I^{(j)}$ and binary foreground $B^{(j)}$ from the anterior and posterior camera modules, for $j <= i$. Two binary images corresponding to the same sliding section of the belt are thus compared, where for each $B_p^{(i)}$ extracted at i, $B_a^{(j)}$ is observed, where:

$$j = i - \lfloor \frac{t_s}{t_r} \rfloor \qquad (2)$$

and t_s is the time necessary for an object to pass from the centre of the anterior camera's field of view to the centre of the posterior's. Given the stationary assumption, if any one detected object in B_a is not present within d pixels from the centre of its corresponding location in B_p, it can assumed the human operator has influenced the state of the object, and a label can thus be generated. The d parameter is arbitrary although we argue that, should the stationary assumption hold true, the centre of the object should not move any more than the radius of the circle circumscribing it.

C. Deep Learning System

The Learning Module developed for the system is based on a CNN Architecture re-designed to be able to deal with streaming, "on-line", data. Each data point, then, is always to be considered a new, unseen, sample. Moreover, retraining on all the previously seen samples is impossible, if training and inference are to happen in parallel to the running of the Detection Modules, and the number of data points can grow unconditionally. We introduce and implement three main concepts, to allow the network to deal with the above issues: object framing, batch buffering and sliding window.

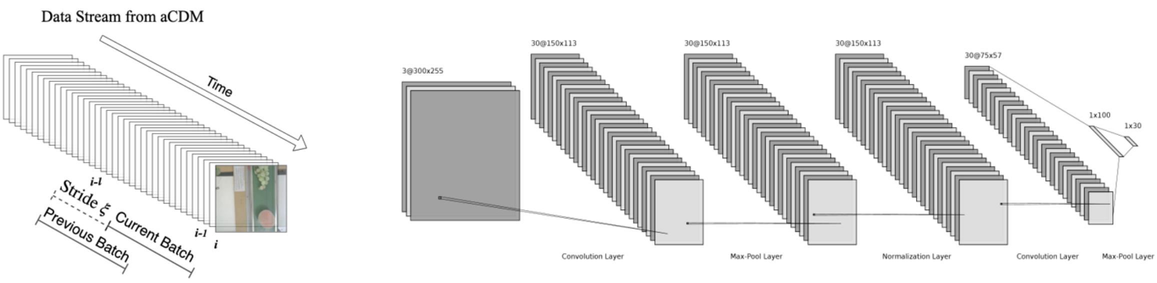
1) *Object framing:* As the aCDM and pCDM modules monitor the conveyor belt, one or more objects can be detected simultaneously in each $I^{(i)}$ at time i. For the CNN to be able to learn the visual features corresponding to each object, it is necessary that each object is: first, separated into a different input; two, made comparable to other objects. To achieve this, for each $I^{(i)}$ the minimum circumscribing boxes containing the binary blobs detected in the

corresponding mask $B^{(i)}$ are found (Fig. 5a, 5b and 5c). The area in the image within the box is then cropped, rotated and padded to re-shape each input array into a comparable size and format (Fig. 5d and 5e).

2) Batch buffering and sliding window: With streaming data, like any other data-set, it is necessary to discourage over-fitting. It is here possible to use a procedure similar to batch training, by buffering the re-shaped objects from $I^{(i)}$ to $I^{(j)}$ for $j = i - \iota$, and training on objects extracted from the batch, rather than from single new images. Parallel to the concept of batch buffering is every how often to train on the buffered images from the past. If batch-training on every single incoming image, the same image will be seen by the network at least ι times, and thus over-fitting is possible for large ι.

We set a stride parameter ξ, which corresponds to the minimum number of time steps i necessary for the network to re-train on the current buffer (Fig. 6a).

3) Architecture and training: The Convolutional Neural Network Architecture designed is composed of two convolutional layers, two pool layers, a normalization layer and two fully connected output layers, as shown in Fig. 6b. All units perform a ReLu non-linear



(a) Data streaming. (b) Convolutional Neural Network architecture.

Figure 6: The CNN architecture and data streaming feed for learning.

(a) aCDM (b) pCDM

Figure 7: The figure an example (a) capture Ia for j = i − and (b) capture . The green boxes correspond to the minimum circumscribing boxes containing the blobs detected in and (Fig. 4c). One of the two detected objects in the aCDM is not present in the poster posterior camera module, thus the object will be marked as influenced by human operations.

transformation [23]. We train the network with the RMSProp adaptive learning rule [24], where the weighted gradient for each weight w at time i is:

$$\Delta w(i) = \frac{-\eta}{\sqrt{S(i)} + \epsilon} \frac{\delta E}{\delta w(i)} \qquad (3)$$

$$S(i) = \beta S(i-1) + (1-\beta)\frac{\delta E}{\delta w(i)}^{2} \quad (4)$$

Here, $S(i)$ corresponds to the weighted average of the square sum of gradients up to i, and is a value to prevent division by zero, set to $1e{-}10$ throughout the experiments. The learning rate η and the decay hyper-parameter β are set to respectively 0.0001 and 0.9, two known good values for the adaptive learning rule [24]. The error is computed as a regularized softmax cross-entropy with logits on the classes absence (0) versus presence (1) of the object in the posterior camera's view.

Results

To test the developed framework the belt system is run, together with the aCDM, pCDM, OLM and Learning module concurrently, on a set of unknown objects. The chosen objects resemble fruit items and are shown in Fig. 8. Three separate tests are performed: object detection, autonomous on-line learning and framework analysis. The performed tests consist in the continuous operation of the conveyor belt described in Section III-B1, while placing the objects in Fig. 8 in line in the anterior's camera's field of view, selectively picking them in the human operational area, and after reaching the end of the belt finally removing and reinserting the objects in the belt, in various positions and orientations (see Video attachment).

A. Object detection

To test the ability of the aCDM and pCDM modules to detect objects The conveyor belt system is run for a total of 15 minutes. For each incoming frame, the binary mask is computed as described previously, and the minimal circumscribing boxes are detected (Fig. 7).

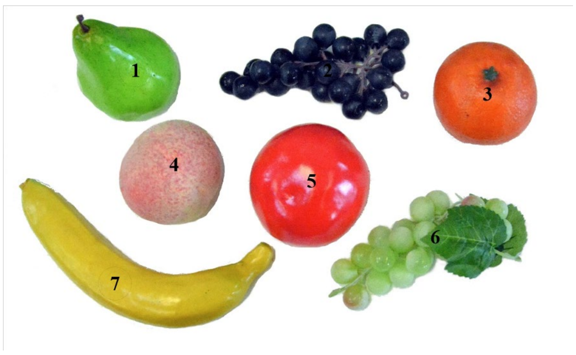


Figure 8: The objects used for the experiments.

aCDM Correct Detections (TP+T
aCDM Incorrect Detections (FP+FN
pCDM Correct Detections (TP+T
pCDM Incorrect Detections (FP+FN

No. of frames	3008 frames
Accuracy	97.67%
pCDM Accuracy	97.63%

TABLE I: The Table reports the detections when testing the system on ≈ 15 minutes of conveyor belt operation.

By comparing the detected items in the aCDM (Fig. 7a) and pCDM (Fig. 7b) it is possible to identify which items have been removed in the human operational area.

After manually labelling the aCDM and pCDM recordings, it is possible to rate the accuracy of the detection system in detecting all objects in the aCDM and corresponding missing objects in the pCDM (Table I). An object is successfully recognized if the distance between its detected and labeled centre is within a range smaller or equal to the radius of the circle circumscribing the object.

As shown in Table I, the system is capable of detecting items with over 97% accuracy over 4187 objects processed within 3008 frames. The mis-detections where mainly due to the GMM object tracker, at times misled by largely varying light during testing. For each object detected by the aCDM, however, it was always possible to detect its presence or absence at posteriori, after human intervention.

B. Autonomous On-Line Learning and Parameter Tuning

To investigate the influence of both the batch buffer size ι, and the stride parameter ξ to the learning, the conveyor system is operated for 15 minutes, and the task is set to be the removal of object 2 (Fig 8). While data is being streamed from the aCDM and pCDM modules, the learning systems learns to recognize those objects which are removed in the Human Operational Area.

Figure 9 shows the inference error and accuracy observed during the conveyor belt operation in the experiment, when learning with varying buffer sizes. The buffer size parameter has a strong influence in the ability for the learning framework to prevent over-fitting, with larger buffers preventing over-fits to the last seen image captures, and thus inducing more stable learning. On the other hand, larger buffer sizes does not allow fast re-learning to support online streams. We set $\iota = 100$, since no noticeable differences were observed in the learning curve for buffers larger than 100 processed frames (Fig. 9). Figure 10 shows 3 learning curves when learning with varying window strides. The stride has an effect both on the number of times the same frame is seen by the network during training, and the speed of learning. Training for every unseen new frame from the aCDM is undesirable, since for a large ι over-fitting is likely. Too large strides might not allow the network to change its weights enough to account for unseen frames. We pick $\xi = 5$, reaching the lowest error during the experiments (Fig. 10).

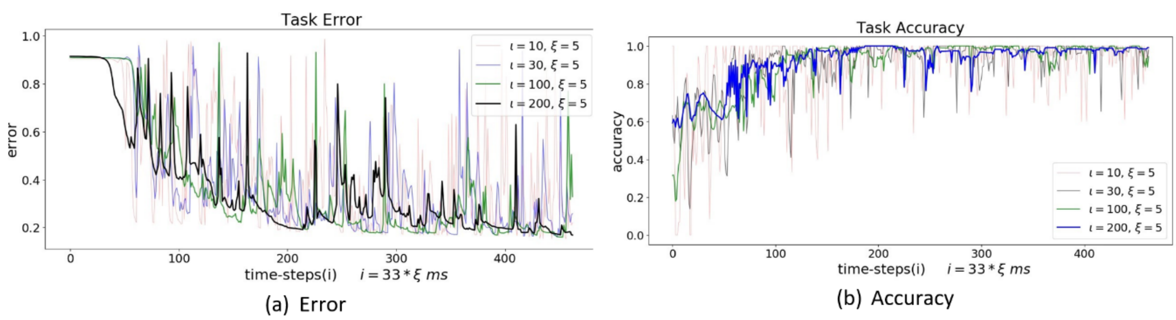


Figure 9: The figure compares the moving (a) error and (b) accuracy when testing the framework with different batch buffer lengths ι. Larger ι values prevent over fitting, thus resulting in more stable learning.

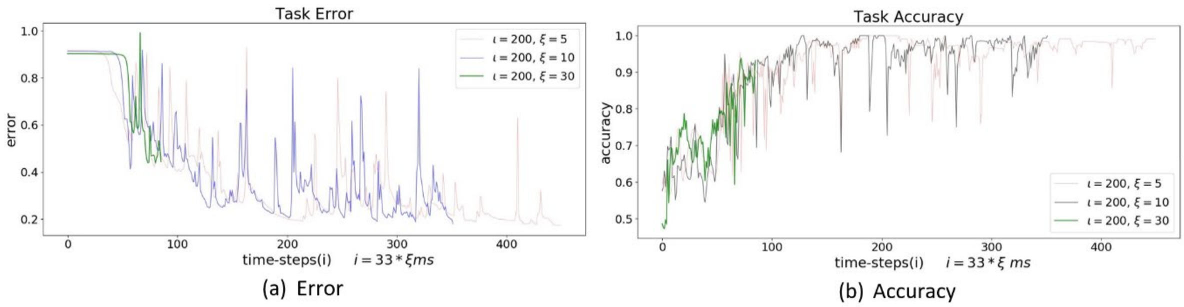


Figure 10: The figure compares the moving (a) error and (b) accuracy when testing the framework with different stride values ξ. Lower values show higher longer training curves.

TABLE II: Test error and accuracy during 5 minutes of continuous streams, and after 10 minutes of on-line training.

C. Framework Analysis

To thoroughly test the framework three separate test runs are performed, each on a separate task: one, removal of the object 1, two removal of object 2 and three removal of object 7 (Fig. 8).

We use the best performing batch buffer and window stride sizes validated in the previous sections, i.e. $\iota = 100$ and $\xi = 5$, and test the accuracy of the network during 5 minutes of continuous streams, and after 10 minutes of on-line training. The errors and accuracies are reported in Table II.

The features relative to object 7 were easier to learn than object 1 or 2, as shown by the higher testing accuracy and error in Table II. The framework reaches stable top performance after ≈ 250 time steps, equivalent to less than 5 minutes of belt operation. The framework is shown to be capable of selectively detecting and identifying objects, both influenced by human operations and not, with an average accuracy of over 97% on all tasks.

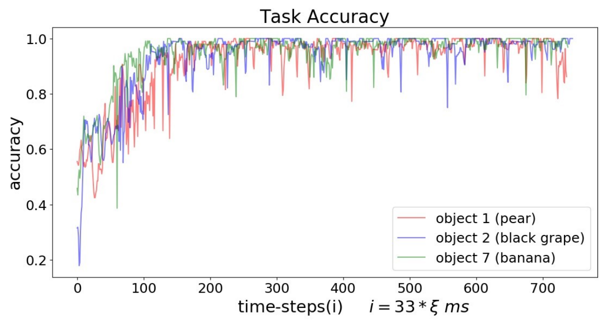


Figure 11: Running accuracy of the framework during testing.

Objects	Error	Accuracy (%)
1 (pear)	0.2383	97.26
2 (black grape)	0.2124	97.93
7 (banana)	0.1966	98.58

TABLE II: Test error and accuracy during 5 minutes of continuous streams, and after 10 minutes of on-line training.

Conclusions

We develop a novel framework to selectively detect and recognize salient objects within a conveyor belt. The framework is capable of learning through observing human labour, and thus needs no explicit data-gathering and training. The adaptability of the framework to various conditions is shown, by purposefully applying object recognition and learning solutions which are feature independent, and thus transferable. We test the framework on a set of unknown objects, which are placed on a custom-made conveyor belt, and selectively picked by a human operator. The system based on the proposed framework is capable of observing human operations and autonomously learning which objects need to be acted upon prior to reaching the human operational area.

Given the GMM-based object detection, object clutter is currently unsolved. Future work will be aimed at augmenting the aCDM and pCDM modules to detect objects in the presence of clutter (or object overlap).

The framework is developed for ease of integration in existing industry environment. Conditional to some basic assumptions, the learning and vision system can run in the background and learn task-specific selective item detection and identification through observation of skilled labour in new environments. No labelling is necessary for the training of the system, and the supervision is supplied seamlessly by the labour, performing the usual required tasks. This work is a step forward toward the full automation of conveyor belt-based systems in non human-friendly environments.

Knowledge and Technology Transfer

- L. Scimeca, F. Iida, "Self-Supervised Learning Through Scene Observation for Selective Item Identification in Conveyor Belt Applications", in Annual Conference Towards Autonomous Robotic Systems (TAROS 2020) (forthcoming), Springer, Cham

- Scimeca, L., Maiolino, P., & Iida, F. (2018, April). Soft morphological processing of tactile stimuli for autonomous category formation. In 2018 IEEE International Conference on Soft Robotics (RoboSoft) (pp. 356-361). IEEE.

- Scimeca, L., Maiolino, P., Cardin-Catalan, D., del Pobil, A. P., Morales, A., & Iida, F. (2019, May). Non-destructive robotic assessment of mango ripeness via multi-point soft haptics. In 2019 International Conference on Robotics and Automation (ICRA) (pp. 1821-1826). IEEE.

- Hughes, J., Gilday, K., Scimeca, L., Garg, S., & Iida, F. (2020). Flexible, adaptive industrial assembly: driving innovation through competition. Intelligent Service Robotics, 13(1), 169-178.

- Hughes, J., Scimeca, L., Ifrim, I., Maiolino, P., & Iida, F. (2018). Achieving robotically peeled lettuce. IEEE Robotics and Automation Letters, 3(4), 4337-4342.

References

[1] R. K. Ranga, S. Yadav, A. Yadav, N. Yadav, S. B. Ranga, *et al.*, "Prevalence of occupational noise induced hearing loss in industrial workers," *Indian Journal of Otology*, vol. 20, no. 3, p. 115, 2014.

[2] H. Piedrahita, J. Oksa, C. Malm, and H. Rintamaki, "Health problems¨ related to working in extreme cold conditions indoors," *International journal of circumpolar Health*, vol. 67, no. 2-3, pp. 279–287, 2008.

[3] R. Inaba, S. M. Mirbod, J. Kurokawa, M. Inoue, and H. Iwata, "Subjective symptoms among female workers and winter working conditions in a consumer cooperative," *Journal of occupational health*, vol. 47, no. 5, pp. 454–465, 2005.

[4] Z. Farhana, S. M. Ali, and M. Rahman, "Investigation on workplace environment and safety–a case study in rahimafrooz batteries ltd." *International Journal of Quality and Innovation*, vol. 1, no. 4, pp. 338– 347, 2011.

[5] P. Maguina, T. L. Palmieri, and D. G. Greenhalgh, "Treadmills: a˜ preventable source of pediatric friction burn injuries," *Journal of Burn Care & Rehabilitation*, vol. 25, no. 2, pp. 201–204, 2004.

[6] R. Krishnamoorthy and G. Karthikeyan, "Degloving injuries of the hand," *Indian journal of plastic surgery: official publication of the Association of Plastic Surgeons of India*, vol. 44, no. 2, p. 227, 2011.

[7] T. Ruff, P. Coleman, and L. Martini, "Machine-related injuries in the us mining industry and priorities for safety research," *International journal of injury control and safety promotion*, vol. 18, no. 1, pp. 11–20, 2011.

[8] S. W. Holland, L. Rossol, and M. R. Ward, "Consight-i: a visioncontrolled robot system for transferring parts from belt conveyors," in *Computer vision and sensor-based robots*. Springer, 1979, pp. 81–100.

[9] M. L. Baird, "Image segmentation technique for locating automotive parts on belt conveyors." in *IJCAI*, 1977, pp. 694–695.

[10] T. H. Park and B. H. Lee, "An approach to robot motion analysis and planning for conveyor tracking," *IEEE transactions on systems, man, and cybernetics*, vol. 22, no. 2, pp. 378–384, 1992.

[11] F. Chaumette, "Visual servoing," *Computer Vision: A Reference Guide*, pp. 869–874, 2014.

[12] N. R. Gans and S. A. Hutchinson, "Stable visual servoing through hybrid switched-system control," *IEEE Transactions on Robotics*, vol. 23, no. 3, pp. 530–540, 2007.

[13] T. Borangiu, F. Anton, and A. Dogar, "Visual robot guidance in conveyor tracking with belt variables," in *2010 IEEE International Conference on Automation, Quality and Testing, Robotics (AQTR)*, vol. 1. IEEE, 2010, pp. 1–6.

[14] R. C. Luo, S.-C. Chou, X.-Y. Yang, and N. Peng, "Hybrid eye-to-hand and eye-in-hand visual servo system for parallel robot conveyor object tracking and fetching," in *IECON 2014-40th Annual Conference of the IEEE Industrial Electronics Society*. IEEE, 2014, pp. 2558–2563.

[15] G. Carbone, *Grasping in robotics*. Springer, 2012, vol. 10.

[16] J. Hughes, U. Culha, F. Giardina, F. Guenther, A. Rosendo, and F. Iida, "Soft manipulators and grippers: a review," *Frontiers in Robotics and AI*, vol. 3, p. 69, 2016.

[17] L. Chin, J. Lipton, M. C. Yuen, R. Kramer-Bottiglio, and D. Rus, "Automated recycling separation enabled by soft robotic material classification," in *2019 2nd IEEE International Conference on Soft Robotics (RoboSoft)*. IEEE, 2019, pp. 102–107.

[18] J. Hughes, L. Scimeca, I. Ifrim, P. Maiolino, and F. Iida, "Achieving robotically peeled lettuce," *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 4337–4342, 2018.

[19] L. Scimeca, C.-C. Daniel, P. Angel, M. Antonio, F. Iida, *et al.*, "Nondestructive robotic assessment of mango ripeness via multi-point soft haptics," 2019.

[20] Z. Zivkovic *et al.*, "Improved adaptive gaussian mixture model for background subtraction." in *ICPR (2)*. Citeseer, 2004, pp. 28–31.

[21] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2000.

[22] Z. Zivkovic and F. Van Der Heijden, "Efficient adaptive density estimation per image pixel for the task of background subtraction," *Pattern recognition letters*, vol. 27, no. 7, pp. 773–780, 2006.

[23] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the 27th international conference on machine learning (ICML-10)*, 2010, pp. 807–814.

[24] T. Tieleman and G. Hinton, "Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude," *COURSERA: Neural networks for machine learning*, vol. 4, no. 2, pp. 26–31, 2012.

