Real-time Barcode Detection in the Wild

Clement Creusot Asim Munawar
IBM Research - Tokyo
{clement,asim}@jp.ibm.com

Abstract

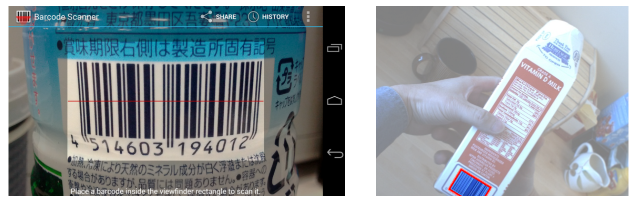
The linear 1D barcode is the main tagging system for billions of products sold each day. Barcodes have many advantages but require a laser scanner for fast and robust scanning. Solutions exist to read barcodes from cellphones but they assume a carefully framed image within the field of view. This undermines the true potential of barcodes in a wide range of scenarios. In this paper we present a real-time technique to detect barcodes in the wild from video streams. Our technique outperforms the state-of-the-art passive techniques both in accuracy and speed. Potential commercial applications enabled by such passive scanning system are also discussed in this paper.

1. Introduction

Linear 1D barcodes appeared in the 1960s and are now present on the packaging of almost all consumer products. A barcode consists of a set of dark stripes on a light background encoding an identification number or ID (see Figure 1-a). Many more advanced tagging technologies have emerged in the last decades allowing the storage of more information (e.g. 2D barcodes, QR codes, RadioFrequency ID (RFID) and so on). However, none of them has really taken off in the retail industry as most applications only require a simple label (for billing, tracking, counting, etc). Barcoding is a very cheap and reliable way to provide universal tagging of products and is likely to remain the de facto industry standard for the coming decades.

The primary strength of the barcode is that it can be read quickly and robustly using a laser scanner. The scanners are often attached or embedded in cash registers but can also be mobile (e.g. for inventory or cash-register-free retail). In both cases the laser beam need to be close to the barcode and is often brought into physical contact.

Solutions exist to read barcodes using only passive 2D vision (without a laser scanner), and are popular apps for smartphones. However current techniques have severe flaws that limit their usability, and therefore, their use. The main difficulty encountered with this type of system is the need for clean, well framed, close to target pictures. Current techniques expect the input to look more like flat bed scans than photographs. This makes the scanning painfully slow and prone to error while requiring the focused cooperation of the user. Figure 1 shows the difference of the inputs between the existing passive barcode scanners and our targeted type of application.



a- Android Barcode Scanner b- Our Detector

Figure 1. Difference between existing and targeted capabilities: a- Mobile apps reading barcodes require a very clean framed input, b- barcodes in the wild should be captured without the user intervention.

Both 2D vision and laser-based approaches require the barcode to be targeted by the sensor. Both approaches would generally fail to find a barcode in a general scene. They assume a human operator has already detected the barcode and will show them that particular region of interest. In other words, existing techniques only "read" the barcodes, they do not "detect" them in a large image. This task is always left to the human operator who brings the barcodes to the readers (supermarket cash register) or the readers to the barcodes (hand held scanners or smartphone barcode reader).

What would be ideal is a system that allows detecting and scaning barcodes smoothly in real-time without the need for user interaction. As soon as a barcode appears in the field of view of, for example, a wearable camera, the device will know what object is being looked at. This would enable a large number of possible consumer applications and customer-seller interactions.

In this paper we propose a new 2D real-time technique to detect barcodes in the wild. We evaluated our technique on several open datasets and demonstrate they outperform existing competing systems. In particular, we found an average increase of 14% in detection accuracy over the state-of-the-art method and a reduction of roughly 20% for the computation time.

In the following section, we will discuss previous approaches proposed to solve this problem and highlight their limitations. We will then present our new technique and evaluate it on two publicly available datasets. A presentation of possible applications enabled by our technique is given before proposing our plans for future development.

2. Previous work

Existing techniques for barcode detection using passive vision can roughly be divided into four groups based on the image processing techniques used. These groups contain the techniques based on morphological operations, image scanning, bottom-hat filtering, and distance transforms. Some of these groups are discussed in detail by Katona et al. [6].

Katona et al. [6] discuss a morphological operations based technique to detect barcodes. In another paper [7] they present a technique for 1D and 2D barcode detection by exploiting the properties of barcodes such as the almost regular spacing of parallel stripes. The technique is suitable for barcodes of known sizes but due to primitive operators like erosion, dilation, and area thresholding, it tends to be unsuitable for images where the size of the barcode is unknown. Other approaches like Juett and Qi [5] also use primitive operators such as thresholding and erosion. These operators are highly dependent on the size and colors of the objects in the image. Finding the optimal parameters for such operators is not always easy. In addition, this work cannot find multiple barcodes in an image.

Bodnar et al. [3] propose a technique to find the barcodes by exploiting their symmetry. This technique falls into the image scanning category. In this method a circle is overlaid on the image to obtain a one-dimensional profile of zero-crossings with various densities. The profile is then divided into ”wild” and ”calm” quadrants to determine the presence or absence of barcodes. In addition to the sensitivity to the tile size selection, this algorithm fails to produce correct results if some of the neighboring barcode have a variable width.

Another recent project by Bodnar et al. [2] exploits the similar properties of tiles of the images for 1D and 2D barcodes in distance transform domain. The areas that show similar properties are clustered and assigned a higher probability of being a barcode. Although suitable for 1D and 2D barcodes, this approach uses edge detection similar to Canny, for which selecting the optimal parameters is a difficult task. Also, the algorithm has a tolerance value that can be adjusted to control the compromise between speed and accuracy. In another approach by Bodnar et al. [1] they attempt to combine simple detectors using various aggregation methods to improve the detection accuracy. As each algorithm involved in the detection cost computational resources, the aggregated time required for this kind of approach is well above the thresholds for real-time processing.

Most of the papers we mentioned evaluate their performance on synthetic or home-made non-public datasets. This makes it difficult to compare results. The earliest proper evaluation we found with publicly accessible datasets was proposed in Zamberletti et al. [12]. They used machine learning techniques to identify 1D barcodes directly in the Hough transform space. Due to the robustness introduced by machine learning the technique is also capable of detection partially occluded barcodes. As far as we know this represents the state-of-the-art barcode detection method and we compare our results against theirs.

3. Proposed Method

As mentioned in Section 2, most existing techniques look at barcodes by either considering them as lines (Hough-transform-based techniques) or textures (morphological techniques). In this paper we use a different approach based on detecting, filtering and clustering blobs. Most specifically, we take advantage of the detection stability of the Maximal Stable Extremal Region (MSER) system introduced by [8]. This allows us to avoid the computation of gradient-based intermediate results such as Canny edge detection to improve the stability and limit the number of parameters. The shape of the blobs is a strong first clue for filtering. The barcode regions are detected by clustering in a transformed feature space that uses, among other things, a representation of the perpendicular middle line to the barcode (which is the same for every blob of the same barcode).

We present the steps of our method in order. Examples of intermediary results for each step can be seen in Figure 2.

3.1. Dark Bar detection using MSER

In 2002, Matas et al [8] proposed an efficient method to compute a stable segmentation of regions for stereo matching. The technique ends up selecting regions that are stable by thresholding over a wide range of intensity values. This is very convenient since it means the thresholding does not need to be global. The boundary of selected regions will follow different intensity values in different part of the image. The only essential parameter to provide is the size of the threshold gap over which a region should be stable. In this paper a region is considered stable if it does not change for 5 consecutive thresholds within the intensity range 0-256. This blob detection system has a complexity of $\mathcal{O}(n\ \log(\log(n)))$ and runs quite fast in our tests, making it a good candidate for real-time processing.

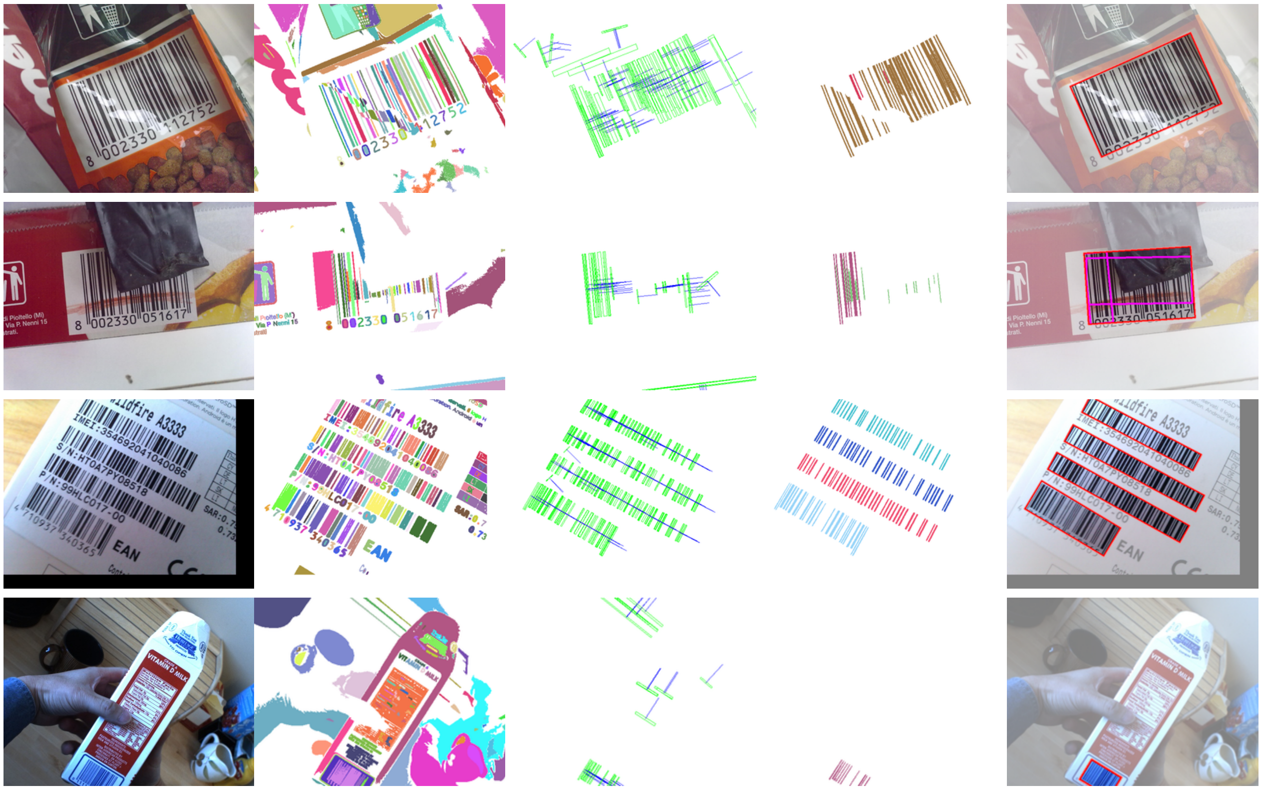


Figure 2. Workflow of the system on a few examples from the ArTe-Lab Rotated Dataset [12] (row 1-3) and from the Intel Egocentric Object Recognition dataset [10] (row 4) where a barcode appears accidentally in the field of view. The segmentation colors are randomly generated and have no special meaning. In the third column, the blobs are represented by their bounding rectangle in green and their perpendicular direction in blue. In the last column, the intermediary detection are in purple and the final one in red.

3.2. Candidate filtering

The MSER technique will segment a huge number of regions within the images. We start by eliminating the obvious false positives based on the aspect ratio of each blob. First, an optimally oriented rectangle including the blob is computed and defined by its center C_i, its with w_i, its height h_i and its orientation angle α_i. Note that if the height is smaller than the width we swap them and modify the angle by 90 degrees.

A detected blob is kept as a candidate if the height to width ratio is above 10. This is a loose threshold to include even very thin truncated barcodes in which the individual black bars are proportionally larger. For a UPC-A barcode, as an example, this ratio will usually be above 24 for non truncated tags and above 14 for truncated ones.

Many people have used the Hough transform approach for barcode detection. However this is often used to detect the many vertical black bars of the barcode using the pixels as evidence for all possible lines in a histogram. Here we use the Hough transform idea to detect a very different line: the imaginary line that is perpendicular to the barcode and passes through the center of each bar. Because of the duality between lines and points in 2D, this line can be presented as a single point relative to a fix origin. This point is the closest point to the origin along the line in Cartesian space and can be represented by its polar coordinates in the Hough space. Since all of the black bars of the barcode are ideally parallel within a rectangle, the bisector of each segment should be the same line and should therefore cluster in a single point in the Hough space. In practice the bisector will vary from segment to segment but still remain similar enough to allow the clustering of segments into barcodes.

The angle and the length of the blob rectangle are also used in the feature vector to discriminate between the blobs from different barcodes. In this step we use a very simple cutting tree approach to clustering. This only needs two parameters: the maximal distance in the feature space between any two branches in the tree and the minimum number of items per branch to be considered a potential barcode. Here again we choose a very loose threshold, since only 5 detections are required for a group of parallel lines to be defined as candidate.

Datasets:	ArTe-Lab [12]	Muenster [11]
#Images	365	1055
#Ground-Truth	365	595
Dimensions	640x480	640x480*
Device	Nokia 5800 + others	Nokia N95

* Other dimensions are available up to 2592x1944.

Table 1. Datasets details

3.3. Post-processing

In practice, the steps already discribed do a good job of detecting barcodes. However some errors still occur in a few cases. One example is when the barcode is parallel with the edge of the box (which is usual). In this case, the border of the box tends to be consider as an extra bar in the barcode. In that particular case, the global direction and center line of the barcode have been correctly found but the beginning and end along that direction are still unclear. To address this problem, we look at the projected histogram along the retrieved bisector of the barcode. If a gap between two spikes is larger than a third of the size of the detected barcode, then the detection is split into two parts on each side of this division. In the last stage we also merge the detected rectangles that intersect each other and use the circumscribed rectangle for both areas as a new detection.

4. Evaluation

In this section we present the evaluation protocol and results on two publicly available datasets. We compare our results with the method proposed in Zamberletti et al [12].

4.1. Dataset

There are not many annotated datasets for barcode detection evaluation. We compare the performances on the datasets used in [12] (see Table 1). The first dataset is the ArTe-Lab Rotated Barcode Dataset (extended version)[1] by [12]. It consists of 365 images containing EAN barcodes captured with different phones. The results given in [12] seems to use only a subset of 129 of those images for the evaluations since they use the rest for training. Here we compare the results using the full set of images. The ground truth they provided is given as a binary mask for each image.

The second dataset is the WWU Muenster Barcode Database[2] from [11]. It contains 1,055 pictures of EAN and UPC-A barcodes taken with a N95 mobile phone. The ground truth is not available for the whole dataset. Ground truth binary masks are provided for 595 of these images in [12].

[1] http://artelab.dista.uninsubria.it/download
[2] http://cvpr.uni-muenster.de/research/barcode/Database/

Datasets	Accuracy $J_{\text{avg.}}$		Detection Rate $D_{0.5}$	
	Zamberletti	Ours	Zamberletti	Ours
ArTe-Lab	0.695	**0.763**	0.805	**0.893**
Muenster	0.682	**0.799**	0.829	**0.963**

Table 2. Detection results on the two datasets.

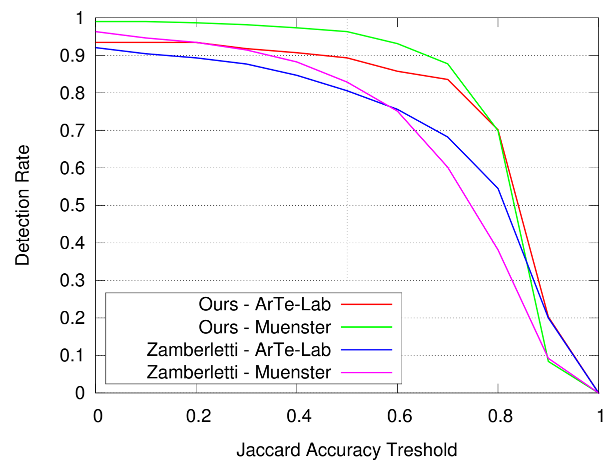


Figure 3. Comparison of detection rates on the two datasets.

4.2. Metrics

To simplify the comparisons we use the same metric as [12]. The main measure aims at evaluating the overall bounding box detection accuracy using the Jaccard index between the ground truth and the detection. Both the detection result R and the ground truth G are given as binary masks over the whole image.

$$J(R, G) = \frac{|R \cap G|}{|R \cup G|}$$

We call $J_{\text{avg.}}$ the average Jaccard accuracy over the dataset. The overall detection rate $D_{0.5}$ corresponds to the proportion of the files in the dataset achieving at least 0.5 Jaccard accuracy (also called OA^{bb} in [12]):

$$D_{0.5} = \frac{\#(i \in S \,|\, J(R_i, G_i) \geq 0.5)}{|S|}$$

with S the set of files in the dataset. For completeness, we provide the detection for varying accuracy thresholds in steps of 0.1. The results for the two datasets can be seen in Figure 3.

4.3. Results

The comparisons of the detection accuracy between our method and [12] can be seen in Table 2. For a Jaccard threshold of 0.5 we succeeded in detecting the barcodes in 89.3% of the images for the ArTe-Lab Rotated dataset

Detection (red) MSERs

Figure 5. Examples of failures caused by blur (rows 1 and 2) and non-compliance to the standard (row 3) with no white margin on edge of the barcode.

Datasets	Zamberletti	Ours
ArTe-Lab	0.135	**0.105**
Muenster	0.130	**0.115**

Table 3. Average computation time in seconds.

and in 96.3% for the Muenster dataset compared to 80.5% and 82.9%, respectively, with the reference technique. The complete detection profile can be seen for the two methods on both datasets in Figure 3. The images where failures occur are usually very blurry, as can be seen in Figure 5. In Figure 4, some examples of our detection (in green) are compared with the ground truth (in black) and the detection from [12] (in red). In many cases, the latest method is confused by straight lines that do not correspond to barcodes. It should be noted that the ArTe-Lab rotated dataset contains only one ground truth mask per image even if several barcodes are present, as in the final two examples of Figure 4. Some of our False Positives are actually True Positives. On average over the two datasets, our approach represents an increase in detection of about 14% over the state-of-the-art.

4.4. Real-time aspect

Our method runs at around 10 frames per second on 640x480 images. In Table 3, we compare our performances with Zamberletti et al [12] while running with a single thread on the same Dell XPS 9100 machine. Both methods are implemented in C++ and use the OpenCV Library. The most computationaly expensive part is the MSERs computation that take an average of 49 ms per frame. The filtering, feature detection, and clustering take respectively 18, 6, and 25 milliseconds. Please note that only the MSERs detection computation depends directly on the size of the image. The other parts depend on the number of candidate blobs. Therefore the computation of the last three steps should not increase much for higher resolution images. On 640x480 images our method contributes to a speed-up of around 1.2 over the reference method. A live mode of our program using a webcam video stream in input demonstrates the realtime capability of our method.

4.5. Limitations

One obvious limitation of our approach is that it is unable to deal with large amounts of blur as seen in Figure 5. The MSERs regions can only be detected if there are some rough boundaries where the regions remain stable for several intensity thresholds. However in that case the blur problem does not affect only the detectability. This leads us to the second main problem: A barcode that is detected and cleanly cropped is not necessarly readable with current decoding techniques. In a small experiment we used the most popular barcode reading library ZXing[3] (C++ version) on barcodes cleanly cropped using the ground truth of the ArTe-Lab dataset. The software was only able to read 77.99% of the clean barcodes. We think this can and should be improved as a direction for future work.

5. Applications

Linear barcoding is clearly a technology of the past, but barcodes are everywhere and will likely remain the de facto industry standard tagging system for a very long time. In this section we show that combining this old technology with newer computer vision capabilities can enable many new applications. Here we focus on applications in human-friendly environments such as shops and supermarkets, where the consumer does not have special scanners except for the consumers modern wearable sensing devices.

In the following scenarios we assume that the human participant is wearing a camera that captures images of what the person is seeing. This could be either a head-mounted camera, such as a Google Glass [4], or a badge-style lifelogging camera, such as a Narrative Clip [9]. One important criterion is that the camera should not need to be manipulated by the user, but passively sees whatever the user is seeing.

5.1. Applications for Customers

Here we give some examples of hands-free applications for normal consumers in shops.

Product Information for Consumers If your device can seamlessly read the barcodes of the objects you are holding, this could help you obtain a great deal of extra information

[3] https://github.com/zxing/zxing

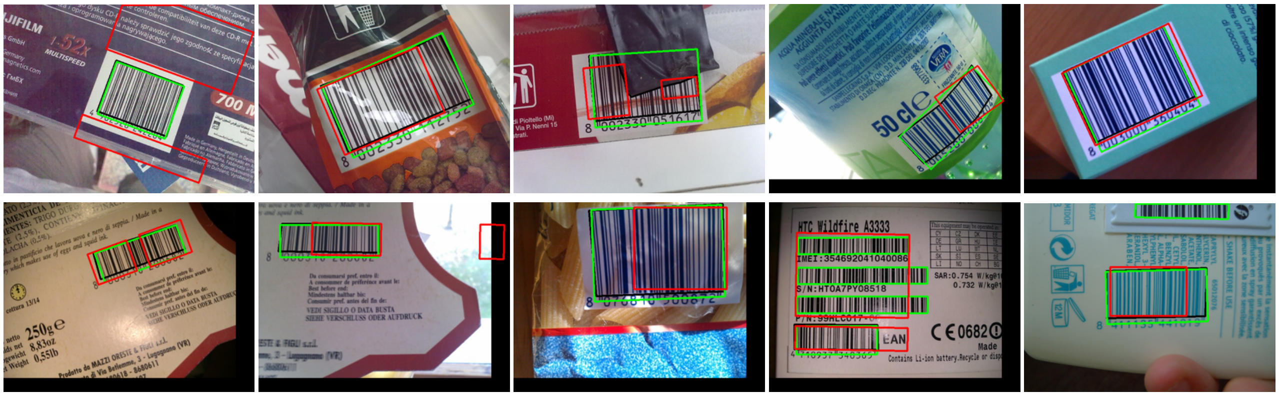


Figure 4. Examples of detection results on the ArTe-Lab Rotated dataset. The ground truth segmentation is in black, the detections from [12] in red and our results are shown in green.

about the products via the Internet. For packaged food, it could tell you the average price, the ingredients, whether it suits your particular dietary requirements (allergies, environmentally friendly, religious beliefs, calorie intake) and so on. For a book or a movie, it could tell you about the reviews, or tell you about the authors and their other books or movies. Obviously this can already be done with a smartphone but requires an active scanning action from the user followed by an active search for information. In the future you might just ask your wearable device "Glass, give me a synopsis of this book", and the device will use the barcode to retrieve the information.

Retailer Information for Consumers If the consumer can only access the Internet the information retrieval will be quite static. However if the shop provides an interface for connected devices to query for information, then you will be able to find out without any shop assistante if one particular garment exists in other colors, whether the current shop has that item in your size, and the exact price in this shop. The retailer can also recommend other products to you or show targeted ads in the shop depending on your apparent interests.

Consumer Feedback The consumer opinions about products are of great value to improve product quality, adjust prices and marketing campaigns, and advise other consumers. With wearable devices there are several ways a consumer can give feedback about the product he or she is seeing:

- By giving private feedback: "Glass, tell the retailer this t-shirt is damaged" or "Glass, tell the manufacturer this product is too fatty".
- By creating public or semi-public reviews: "Glass, Tweet: This is a nice pair of shoes, remember my birthday!"
- By supporting passive anonymous surveys: 100 people looked at this item's price tag today and only one bought it. Therefore people are interested in this article but evidently think it is too expensive.

In each of these cases the identity of the product is central. Having a universal and effortless way to capture this information and add it to our interactions could have an important impact.

Life Management A universal and passive barcode reader could also enable very interesting applications for daily life management and life-logging. A very simple example is shopping list management. By just looking at the back of your cereal box before putting it in your cart the system could automatically erase it from your shopping list and then only remind you about the missing items. This kind of shopping assistant application could help with many other tasks, giving advices, product and price comparisons and so on. Another possible application would be consumption life-logging. By adding the ID of the products you are buying or using to the stream of information you generate, you can increase the level of insight you have on your own life style. You could ask your life-logging database questions about your consumption such as: "How much chocolate did I buy this year?". This could also be valuable in academic research for large scale study on diets, diseases, and so on.

5.2. Limitations and Alternatives

In most cases, the amount of action required by the user is very limited and only occurs if the barcode is not visible. In that case the user needs to rotate the object. This can even been done if the user is visually impaired. When shopping for clothes it usually requires no extra action, since the price and barcode are on the same tag.

One important consideration is that such a system is not a panacea. In theory, it should not be necessary, since humans can normally recognize objects without using barcodes, and they are also not sufficient, since some products (such as fresh food products) have no tags. There are alternative identification methods worth considering. One booming field of research is visual object recognition using the appearance of the objects alone, the way normal humans recognize objects. However there are three obvious limitations. First, the research in this area is really only starting and in spite of the leap forward initiated by deep learning no existing system can pretend to classify accurately the content of, for example, a small supermarket. In contrast, reading barcodes is possible right now. Second, products might change their appearance regularly for marketing reasons making them unsuitable for appearance-based matching (such as a box of cereal with promotional toys). The barcode does not and should not change. Third, even if machine vision achieved human-level object recognition, even the accuracy of human labeling is somewhat limited. A bottle of soda with a big brand might be very easy to recognize and associate with a product name but it is much more difficult for small brands or limited production items, such as a pair of shoes. Being able to precisely access the object ID is very important. It makes all the difference between "I want a pair of green shoes" and "I want *this* pair of green shoes".

6. Conclusion

The multiplication of wearable sensors in our lives is increasing the desire for detection and recognition within non-framed images in the wild. This is of course true for text, which is much more difficult to recognize on a shop window in the street than from a book in a flat-bed scanner. The same applies to barcodes. However we think the task will be solved for barcodes long before it is solved for general text, since their appearance is standardized.

Ultimately, the aim of many computer vision research projects is to extract semantic labels from images. A frequent complaint in this area is the lack of labeled training data. The good news is that we routinely see many barcodes that convey this semantic labeling in our daily lives. By being able to read these accidental barcodes and associate them with tracked objects in video segments, we might be able to generate huge amounts of labeled data for millions of class of objects. Of course this will not help the labeling of cats and dogs, but will be of great use for a lot of practical applications, including assistance robots and vision glasses for the visually impaired and aging populations.

In conclusion, we should acknowledge that barcodes are everywhere. It would be a waste not to use them more effectively.

References

[1] P. Bodnar and L. Nyul. Improving barcode detection with combination of simple detectors. In *Signal Image Technology and Internet Based Systems (SITIS)*, pages 300–306, Nov 2012.

[2] P. Bodnar and L. G. Nyul. Barcode detection with uniform partitioning and distance transformation. In *IASTED International Conference on Computer Graphics and Imaging*, pages 48–53, Feb 2013.

[3] P. Bodnar and L. G. Nyul. A novel method for barcode localization in image domain. In M. Kamel and A. Campilho, editors, *Image Analysis and Recognition*, volume 7950 of *Lecture Notes in Computer Science*, pages 189–196. Springer Berlin Heidelberg, 2013.

[4] Google. Glass, 2014. https://www.google.com/glass/start/.

[5] J. Juett and X. Qi. Barcode localization using a bottom hat filter, 2005.

[6] M. Katona and L. Nyul. A novel method for accurate and efficient barcode detection with morphological operations. In *Signal Image Technology and Internet Based Systems (SITIS)*, pages 307–314, Nov 2012.

[7] M. Katona and L. Nyl. Efficient 1d and 2d barcode detection using mathematical morphology. In *Mathematical Morphology and Its Applications to Signal and Image Processing*, volume 7883 of *Lecture Notes in Computer Science*, pages 464–475. Springer Berlin Heidelberg, 2013.

[8] J. Matas, O. Chum, U. Martin, and T. Pajdla. Robust wide baseline stereo from maximally stable extremal regions. In *Proc. of the British Machine Vision Conference*, pages 384–393, 2002.

[9] Narrative. Narrative clip, 2014. https://getnarrative.com/.

[10] X. Ren and C. Gu. Figure-ground segmentation improves handled object recognition in egocentric video. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 3137–3144, June 2010.

[11] S. Wachenfeld, S. Terlunen, and X. Jiang. Robust recognition of 1-d barcodes using camera phones. In *Proc. of 19th International Conference on Pattern Recognition (ICPR 2008)*, pages 1–4, 2008.

[12] A. Zamberletti, I. Gallo, and S. Albertini. Robust angle invariant 1d barcode detection. In *Pattern Recognition (ACPR), 2013 2nd IAPR Asian Conference on*, pages 160–164, Nov 2013.