Automated Object Tracking in Sterile Pharmacy Compounding

Zhe Qi

Dept. of Computer Science
221 Hayes Hall
Bowling Green State Univ.
Bowling Green, OH 43403,
USA

qizhe@bgsu.edu

Jong Kwan Lee

Dept. of Computer Science
221 Hayes Hall
Bowling Green State Univ.
Bowling Green, OH 43403,
USA

leej@bgsu.edu

ABSTRACT

A new method to automatically track objects in the process of sterile pharmacy compounding is introduced. The method performs the automatic object tracking by applying a simple object detection followed by an object tracking algorithm. The advantages of the method are that it enables both single and multiple object tracking without requiring large sets of templates or training sets. The experimental results show that the method reasonably tracks the objects accurately.

Keywords

Object Detection/Tracking, Image Processing, Pharmacy Compounding Application

1 INTRODUCTION

Object detection and object tracking have been utilized for many application domains (e.g., [**mgi04**, **bps05**, **azl11**]). They play a very important role for the various applications since identifying and tracing the feature(s) of interest are the primary steps for further analysis or new findings. The target application of this paper is the sterile pharmacy compounding.

A majority of medications administered in hospitals, long-term care facilities, patient center medical homes, and emergency medicine are given patients as an injection. Such compounding sterile products are predominantly prepared by mixing, compounding, and manipulating pre-manufactured products in an intravenous compounding room in the hospital or a special compounding center. (Figure 1 shows an example of sterile compounding workspace in a clean room. There is very minimal environmental disruption.) Errors and omissions in this process occur often, leading to severe physical implications for patients (some can be fatal).

Continual monitoring, evaluation, and correction during the compounded sterile product preparation relies

Figure 1: An Example of Sterile Compounding Workspace (image credits given to M. Soper [**sop18**]).

on direct (or indirect) human supervision that is provided by an outside participant or a sole practitioner. There are many pharmacy compounding rules [**gjc13**] to follow, including objects that put on the compounding table should not be close to the table edge; there is a minimum distance between two objects on the table; etc. However, the human factors often result in not following some of the rules and/or lead to errors when compounding products.

In this paper, we introduce a method for tracking objects in sterile pharmacy compounding process. Such a method can assist reducing human errors–it can support the observation of the physician's operations in the real hospital setting as well in the training setting for education. For example, an effective and efficient automated tracking method can signal warnings when human errors occur. Figure 2 shows the simulated sterile compounding workspace setting used in our application.
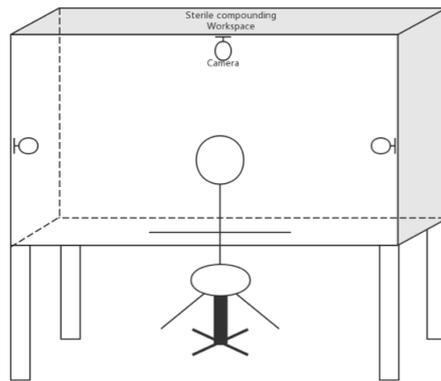
Figure 2: Sterile Compounding Workspace Setting with Cameras.

(While multiple cameras can be used, only one camera was used for this paper.) The video frames taken from the camera can be processed by our method in real time.

The paper is organized as follows. In Section 2, the related work is discussed. The new automated object tracking method for assisting sterile pharmacy compounding is described in Section 3. In Section 4, the experimental results and analysis are presented. Section 5 concludes this paper.

2 RELATED WORK

Many methods for object detection or object tracking in other application domains have been reported in the literature. In this section, some of the recent work are discussed briefly. Here, we note that we intentionally discuss a wide range of literature since our application covers both object detection and object tracking.

Object tracking is often used in intelligent video surveillance and motion recognition in human-computer interaction. One example is the traffic monitoring. Typically traffic monitor employs object tracking methods to identify the traffic-causing vehicle (e.g., accidents) or may predict traffic based on where the vehicles are moving on the road. For example, Rouhani et al. [**rmk17**] have introduced an automated re-configurable framework for real-time intelligent video surveillance. Their framework enables efficient and accurate object tracking by utilizing image processing, principle component analysis, and adaptive learning dictionary matrix for foreground estimation in their intelligence transportation system. Their framework allows efficient processing by employing a field-programmable gate array (FPGA) with accompanying APIs which allow efficient processing.

Panchal et al. [**ppp15**] have reviewed several object detection and tracking methods, including a background subtraction method, real-time background subtraction and shadow detection technique theory, and template matching and shape-based method. Their review included varying environmental settings that can happen in different object recognition and compared the methods. The summary of their review included that (1) the background subtraction was widely used, simple to implement, provided fast recovery with low memory requirement, allowed objects become a part of the background without altering the existing background, but it was not good when shadow and many other objects were present, could not handle very quick objects changes well; (2) template matching was the best method for a specific environment, such as a background invariant environment, but it only supported one-to-one match and worked only if the object appeared in all video frames, it had slow process for recognize new variation of a pattern; (3) frame differencing performed well for static background, it was very easy and gives high accuracy, but it required a background without moving objects.

Recently, Held et al. [**hts16**] introduced a method called GOTURN, which enabled very efficient object tracking using a machine learning-based scheme. Their scheme trained the GOTURN tracker entirely offline with video sequences and images (i.e., no online training was required). It also utilized a local genetic object detector to find target objects and performed frame-by-frame comparisons for tracking. Their method did not include any object labeling and need object type information, which were typically need for other tracker algorithms.

A popular feature detection domain is fingerprint identification and face recognition (e.g., [**olp03**, **kun04**]). The research in this domain has been very active in many decades, but still is one of the important applications. For example, these days, fingerprint-enabled door lock and smartphone fingerprint identification use photoelectric conversion equipment and image processing technique to collect, analyze and compare fingerprints in order to authenticate the users. For example, Kundargi et al. [**kuk18**] recently introduced an automated fingerprint recognition system that overcomes the problem of fake verification caused by an artificial fingerprint from synthetic finger fabricates (e.g., silicone or latex). Their system included a specialized hardware module and a specialized software module for liveness detection, employing a texture feature analysis for verification.

For the work briefly discussed above, most of them utilized a machine learning-based processing. Thus, they needed a large amount of datasets for training. The new automated method introduced in this paper performs object detection and object tracking without a large dataset training process. To our knowledge, no automated method for assisting sterile pharmacy compounding has been presented yet in the literature.
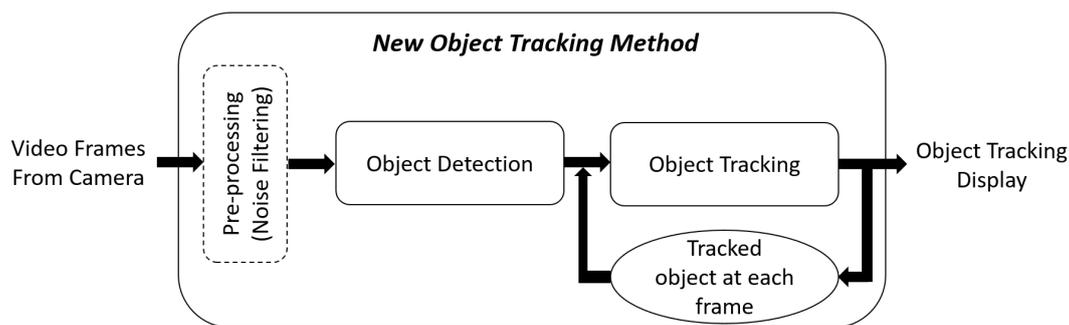
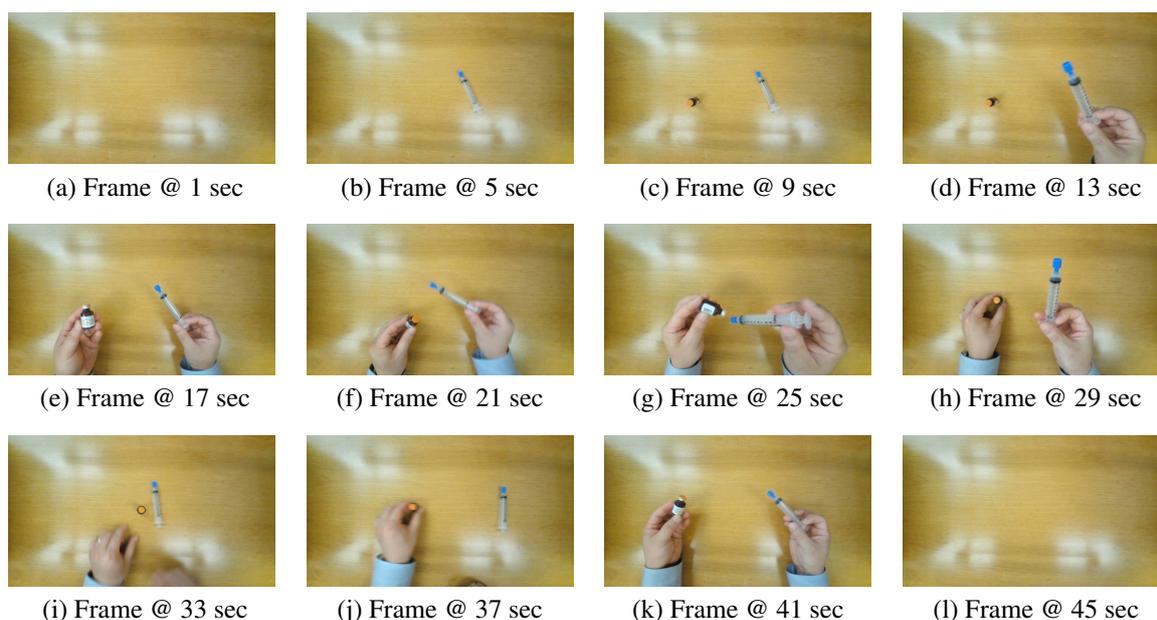Figure 3: New Automated Tracking Method: the Overview.

(a) Frame @ 1 sec   (b) Frame @ 5 sec   (c) Frame @ 9 sec   (d) Frame @ 13 sec

(e) Frame @ 17 sec   (f) Frame @ 21 sec   (g) Frame @ 25 sec   (h) Frame @ 29 sec

(i) Frame @ 33 sec   (j) Frame @ 37 sec   (k) Frame @ 41 sec   (l) Frame @ 45 sec

Figure 4: An Example of Simulated Sterile Compounding Video Frames.

3 NEW AUTOMATED METHOD

Next, the new automated tracking method for assisting sterile pharmacy compounding is described.

The overview of the new method is shown in Figure 3. The image frames of video (i.e., the table scene of sterile compounding) taken from the camera are continuously fed into the tracking steps. The new tracking steps include (1) a pre-processing step to "clean-up" the image frame, (2) the object detection step to initialize the location of objects, and (3) the object tracking step to keep tracking the location of objects.

In our application, two consecutive image frames are considered at each iteration of the steps and the latter frame is re-considered in the next iteration. Figure 4 shows an example of simulated compounding video frames captured at every 4 seconds. (Before subfigures (b) and (c), each object was put on the table by a physician.) As shown in the figures, there are also other image effects (e.g., lights, reflection, shadows) in the

video frames. Here, we note that all videos used in our paper assumed that the table was empty at the beginning (which will be the case in the real setting). Thus, the very first frame is used as the background frame.

3.1 Pre-processing Step

The primary goal of the pre-processing step is to clean up image noise. In our method, a Gaussian filter [**gow02**] is used to de-noise video frames. (We assumed that there would not be complex image noise in the real sterile compounding video frames since it will be taken in a clean room setting. However, we assumed that there would be some normal distributed random noise from video camera sensor.)

3.2 Object Detection Step

After the video frames are cleaned, the object detection step is applied to initialize the locations of objects of

interest. In particular, frame differencing, thresholding, and a morphological closing are applied in the step.

The very first frame in each video is used as the background frame. We denote the background frame as the *B frame*. If a video frame contains any object, the *B* frame can be referenced and compared to find the image regions where the object appears. For example, when a brown bottle appears on frame N, the image regions where the bottle are can be determined by examining the background subtracted image (i.e., $|N - B|$ image and we denote the background subtracted image as the *S* image). When iterating the object detection step with N and $N + 1$ frames, the frame differencing of these consecutive frames gives a local signal of where the objects were moved. We denote the result of this frame differening as the *D image*. In our method, we used a weighted sum of the *S* image and the *D* image to enhance the contrast of the object boundary appeared on the *N* frame.

To locate the object boundary, a simple thresholding is applied (because the image regions with objects will have high value in the object contrast enhanced image) and a contour extraction is applied on the thresholded image. Here, we note that while the physician's hand and arm will also be extracted, we can discard them by removing the larger object contour because they appear as larger contours.

In addition, since light effects (e.g., lights, reflection, shadows) and object occlusion can result in incomplete object contours, a morphological closing is applied to connect the disconnected object regions.

At the end of the detection step, the bounding boxes of the objects are determined by considering the maximum and minimum coordinates of the object contours. Here we note that, the parameters used in this step were determined empirically (e.g., threshold value of 15 and morphological structure of size 5 by 5 were used).

3.3 Object Tracking Step

The last step of the new object tracking method is the object tracking step. In this step, the frame-by-frame objects are being tracked by utilizing a combination of a local template matching and a tracker algorithm.

The cross-correlational template matching [**bru09**] is employed to track the object bounding box on the objects. In particular, the bounding box from the previous frame is used as the template for the current frame. In addition, a tracker algorithm is applied with the template matching for more accurate object tracking by exploring the scale-invariant properties of the tracker algorithm. (For example, the objects in the video often move up and down, resulting in different scales.) Our method utilized two different tracker algorithms: the median flow tracker [**kmm13**] and the discriminative correlation filter-based tracker [**dhf14**]. (We

have chosen these two tracker algorithms since they are used widely and are available on the OpenCV library.) From empirical testings, we have found that the medial flow tracker produced a better result for a single object tracking and the discriminative correlation filter-based tracker produced a better result for a multi-object tracking. In our method, the combination of template matching and tracker algorithm is applied locally (i.e., applied to nearby image regions of the object bounding boxes) since the objects do not typically move very quickly (i.e., there is no abrupt positional change).

The object tracking step is repeatedly applied frame-by-frame to track the objects for the entire video frames.

4 EXPERIMENTAL RESULTS

The effectiveness of the our object tracking method has been benchmarked using the real videos in the simulated workspace setting. Both single and multi-object videos are considered. Each video frame was about 1,920 pixels in width and 1,080 pixels in height and each video included about 1,000 to 1,500 frames. The benchmarking included measuring the method's effectiveness by considering the accuracy and the processing time. In our work, we define the accuracy of the object tracking as the percentage ratio of the number of frames that had at least 75% of the objects detected/tracked in each frame and the number of frames that the objects appeared. We used the average accuracy of all object accuracies for the multi-object tracking benchmarkings. We have compared our method's effectiveness with the effectiveness of the cross-correlational template matching [**bru09**]. All experiments were performed on a PC with Intel Core i7 2.7GHz CPU with a 8GB RAM.

Figure 5 shows the examples of object tracking using our method. Figure 5 (a) and (b) are for a single-object tracking (with a brown bottle of sterile compounding product) and (c) and (d) are for a multi-object tracking (with the brown bottle and a needle tube). The single-object and multi-object tracking videos had 904 frames and 1,368 frames, respectively. The red rectangle overlays in each sub-figures are the markers for object tracking. As shown in the figure, even though there were different light effects, the new method detected and tracked the objects accurately- the single-object and multi-object tracking shown in Figure 5 had about 84% and 80% accuracies, respectively. (Additional discussion for improvement is in Section 5.)

While the new tracking method had the tracker makers on all objects for all frames the objected appeared, there were some cases it did not accurately marked the objects. As defined above, we consider the tracking for a frame is a miss if the method didn't mark at least 75% of the object in each frame. Figure 6 shows two examples where the tracking miss occurred. In Figure 6 (a), while the brown bottle is marked, only about 72%

(a) Single-object Tracking Ex. 1

(b) Single-object Tracking Ex. 2

(c) Multi-object Tracking Ex. 1
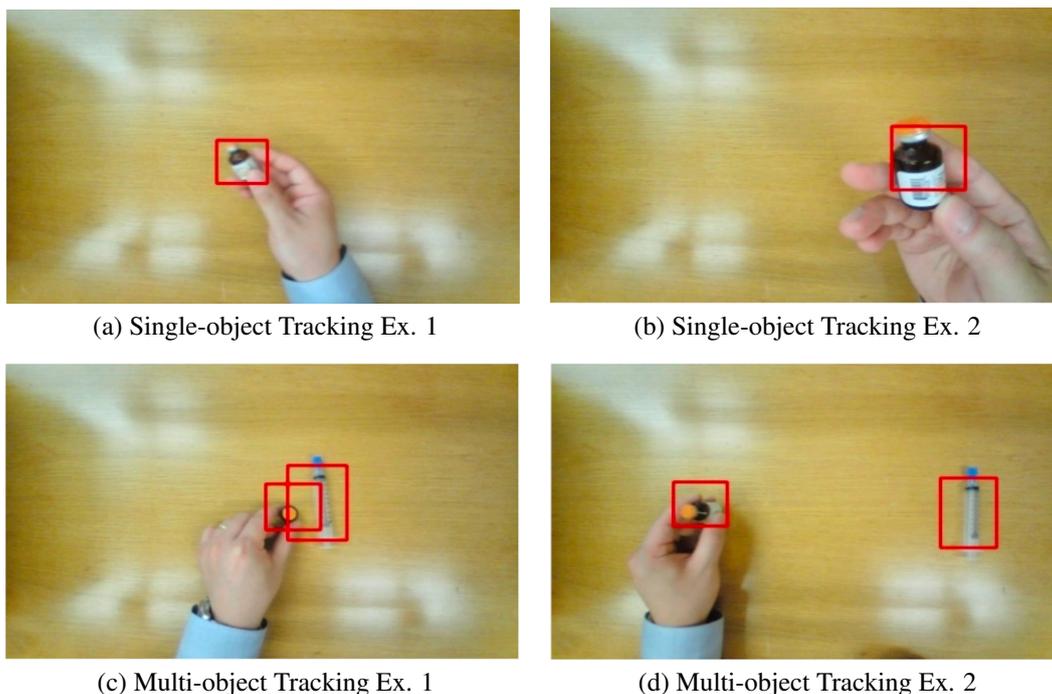

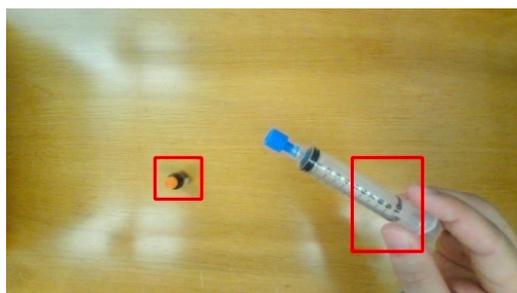(d) Multi-object Tracking Ex. 2

Figure 5: Automated Object Tracking Results: Tracking Frames.

(a) Miss in Single-object Tracking

(b) Miss in Multi-object Tracking

Figure 6: Examples of Missing Objects.

of the bottle was included in the tracking marker. In Figure 6 (b), while the brown bottle is marked well, the new method marked only parts of the needle tubes.

For comparison of methods, the cross-correlational template matching had about 66.51% accuracy and 0.0376 seconds for processing time on average and the new tracking method had about 81.66% accuracy and 0.0415 seconds for processing time on average. While the template matching method was slight more efficient than our method, our method outperformed the accuracy.

5 CONCLUSION

We have presented and evaluated a new object tracking method that can be used in the sterile pharmacy compounding application. The method involves use of a simple image processing, object detection, and object tracking algorithms that enable effective object tracking. One of the primary advantages of the new method is that it does not require a large number templates nor training datasets. The new tracking method produced a very reasonable tracking results and outperformed the the cross-correlational template matching.

For the future work, other image processing steps to identify the objects more accurately may be explored (e.g., a better object contour extraction step, more effective step to remove hand and arm, etc). We also plan to explore additional step to more accurately mark the tracked objects. For example, Hough Transform (HT) and its variation [**hou62**, **hoc96**] may be considered in our application by applying a variant in a local fashion way. (Figure 7 shows our ongoing effort on local HT-based tracking makers as an example. As shown in the figure, the markers are very accurately placed on objects.) Fourier descriptors [**gow02**] may also be
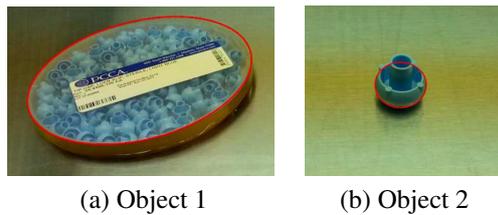
(a) Object 1   (b) Object 2

Figure 7: Accurate Object Labeling with HT.

considered. For efficient processing of the method, we may also explore GPU processing (e.g., [**sim03**, **urg07**, **deg10**]).

We also plans to compare our method with other techniques and methods used for object tracking. (In our preliminary results of comparing with other machine learning-based algorithms, other algorithms need a large number of training datasets to have a reasonable accuracy.)

6 REFERENCES

[AZL11] Aiping W., Zhi-Quan C., and Li M., Multiple-Cue-Based Visual Object Contour Tracking with Incremental Learning, in Proc., 2011 Int'l Conference on Virtual Reality and Visualization, 2011, pp. 30–37.

[BPS05] Bleser G., Pastamov Y., and Stricker D., Real-time 3D Camera Tracking for Industrial Augmented Reality Applications, in Proc., The 13th Int'l Conference in Central Europe on Compute Graphics, Visualization and Computer Vision (WSCG 2005), Jan. 31–Feb. 4, 2005, pp. 47–54.

[Bru09] Brunelli R., Template Matching Techniques in Computer Vision: Theory and Practice, 1st Ed., Wiley, April 2009.

[Deg10] Degirmenci, M., Complex Geometric Primitive Extraction on Graphics Processing Unit, Journal of WSCG, Vol. 18, 2010, pp. 129–134.

[DHF14] Danelljan M., Hager G., Khan F., and Felsberg M., Accurate Scale Estimation for Robust Visual Tracking, in Proc., British Machine Vision Conf., Nottingham, Sep. 2014, pp. 65.1–65.11.

[GJC13] Gudeman J., Jozwiakowski M., Chollet J., and Randell M., Potential Risks of Pharmacy Compounding, Drugs in R&D, Vol. 13 (1), March 2013, pp. 1–8.

[GoW02] Gonzalez, R. and Woods, R., Digital Image Processing, 2nd Ed., Prentice Hall. 2002.

[HTS16] Held D., Thrun S., and Savarese S., Learning to Track at 100 fps with Deep Regression Networks, in Proc., European Conf. on Comp. Vision, Springer, Cham, Oct. 8, 2016, pp. 749–765.

[HoC96] Ho, C. and Chen, L., A High-speed Algorithm for Elliptical Object Detection, IEEE Transaction on Image Processing, Vol. 5 (3), 1996, pp. 547–550.

[Hou62] Hough, P., Method and Means for Recognizing Complex Patterns, U.S. Patent, 3,069,654, 1962.

[KMM10] Kalal Z., Mikolajczyk K., and Matas J., Forward-backward Error: Automatic Detection of Tracking Failures, in Proc., IEEE 20th International Conference on Pattern Recognition (ICPR), Aug. 2010, pp. 2756–2759.

[KuK18] Kundargi J. and Karandikar R.G., Integrating Liveness Detection Technique into Fingerprint Recognition System: A Review of Various Methodologies Based on Texture Features, Progress in Intelligent Computing Techniques: Theory, Practice, and Appls., 2018, pp. 295–305.

[KuN04] Kukharev G. and Nowosielski A., Visitor Identification - Elaborating Real Time Face Recognition System, in Proc., The 12th Int'l Conference in Central Europe on Compute Graphics, Visualization and Computer Vision (WSCG 2004), Feb. 2–6, 2004, pp. 157–164.

[MGI04] Misu T., Gohshi S., Izumi Y., Fujita Y., and Naemura M., Robust Tracking of Athletes Using Multiple Features of Multiple Views, Journal of WSCG, Vol. 12, 2004, pp. 285–292.

[OLP03] Oh S.-K., Lee J.-J., Park C.H., Kim B.-S., and Park K.-H., New Fingerprint Image Enhancement Using Directional Filter Bank, Journal of WSCG, Vol. 11 (1), 2003, pp. 1–8.

[PPP15] Panchal P., Prajapati G., Patel S., Shah H., and Nasriwala J., A Review on Object Detection and Tracking Methods, International Journal for Research in Emerging Science and Technology, Vol. 2 (1), Jan. 2015, pp. 7–12.

[RMK17] Rouhani B.D., Mirhoseini A., and Koushanfar F., Rise: An Automated Framework for Real-time Intelligent Video Surveillance on FPGA, ACM Trans. on Embedded Computing Systems (TECS), Vol. 16 (5s), Sep. 2017, Article No. 158.

[SIM03] Strzodka, R., Ihrke I., and Magnor, M., A Graphics Hardware Implementation of the Generalized Hough for Fast Object Recognition, Scale, and 3D Pose Detection, 12th Int'l Conf. on Image Analysis and Processing, 2003, pp. 188–193.

[Sop18] Soper M., Why Should Compounding Pharmacies Continuously Monitor Particles?, www.setra.com/blog/why-should-compounding-pharmacies-continuously-monitor-particles, May 03, 2018, accessed on Jan. 30, 2019.

[URG07] Ujaldon, M., Ruiz, A., and Guil, N., On the Computation of the Circle Hough Transform by a GPU Rasterizer, Pattern Recog. Letters, Vol 29, (3), 2007, pp. 309–318.