

ALMA MATER STUDIORUM · UNIVERSITÀ DI BOLOGNA

SCHOOL OF ENGINEERING AND ARCHITECTURE

*DEPARTMENT OF ELECTRICAL, ELECTRONIC AND
INFORMATION ENGINEERING "GUGLIELMO MARCONI" (DEI)*

MASTER DEGREE IN AUTOMATION ENGINEERING
0920

GRADUATION THESIS

in

COMPUTER VISION AND IMAGE PROCESSING M

**OBJECT DETECTION IN ROBOT
PICKING APPLICATIONS USING 3D
CAMERAS**

CANDIDATE:
Francesco Baruzzi

SUPERVISOR:
Eminent Prof. **Luigi Di Stefano**

CO-SUPERVISORS:
Eng. **Guido Toccafondi**

Academic year
2020/2021

Session
V

Contents

Introduction	5
1 Intel RealSense D435	6
1.1 Working principle	6
1.2 Main configuration parameters	7
2 Basler blaze 101	15
2.1 Working Principle	15
2.2 Main configuration parameters	17
3 Application target and specifications	23
4 Products qualitative analysis	25
4.1 Basler blaze 101	26
4.1.1 Transparent objects	27
4.1.2 Flowpack objects	31
4.1.3 Local saturations	33
4.1.4 Multiple reflections	34
4.2 Intel RealSense D435	36
4.2.1 Flowpack objects	37
4.2.2 Transparent objects	38
4.3 Nomenclature	39
4.3.1 Algorithms efficiency parameters	41
5 Main algorithms concept	42
5.1 Connecting contours method	42
5.1.1 Basler detail	45
5.2 Background subtraction method	46
5.3 Search of the Object Model	52
5.3.1 Image prefiltering	52
5.3.2 Matching	54
5.4 Derivative Analysis	57
5.5 Common Area Method	63
5.6 Multi-Layer Perceptrons (MLP) Method	66

5.6.1	Working Principle	67
5.6.2	Method Description	70
5.7	Overall considerations	75
6	Results	77
6.1	Conveyor Belt types	78
6.1.1	White lucid plain conveyor belt	78
6.1.2	Dark opaque plain conveyor belt	80
6.1.3	Lucid V section conveyor belt	81
6.2	Main Settings	83
6.2.1	Intel RealSense D435	86
6.2.2	Basler blaze 101	89
6.3	Algorithms performances	90
6.4	Static Results	98
6.4.1	Intel RealSense D435	98
6.4.2	Basler blaze 101	117
6.5	Dynamic Results	121
6.5.1	Intel RealSense D435	123
6.5.2	Basler blaze 101	133
6.6	Some specific cases	143
7	Conclusions	153
7.1	Images Quality	153
7.1.1	Intel RealSense D435	154
7.1.2	Basler blaze 101	154
7.2	Algorithms Designed	155
7.2.1	Basler blaze 101	157
7.2.2	Intel RealSense D435	159
7.3	Final Considerations	161
	Bibliography	164

Introduction

This work relies on analysis and testing of two 3D cameras in robot picking applications. Continuous progress in microelectronics, micro optics and micro technology made 3D cameras affordable and competitive with respect to 2D cameras in common industrial and commercial applications. 3D cameras, in fact, could give advantages in terms of timing and performances in application involving objects normally processed with 2D cameras. These cameras are intrinsically different in the technology used and provide mono/color images, depth maps and point clouds. First camera considered, Intel RealSense D435, is a stereo camera and the second one, Basler blaze 101, is a time of flight camera. First of all both stereo cameras and time of flight cameras working principles are presented along with their main settings. A particular focus is devoted to enhancing eventual advantages of using one technology with respect to the other. It's important to underline that objects examined are mainly syringes and pipettes, so their dimensions are quite small and heterogeneous in terms of material, opacity, colors, textures and transparency. Qualitative considerations will be done to a priori exclude from tests objects which depth maps result difficult to be created and exploited. Then through computer vision algorithms depth information are used to detect graspable objects and to get their positioning and orientations. These algorithms are capable of managing both isolated, touching or overlapping objects, establishing which one can be picked up and which not. Three different algorithms will be deeply introduced and described with a detailed focus on best camera settings to accomplish these tasks. Camera settings are fundamental to get best depth information from cameras also compensating eventual motion blur cases. Several analyses are done with both static and moving products on conveyor belts, varying conveyors colors and shapes. Finally results about the validity and effectiveness of proposed algorithms are illustrated in detail.

Chapter 1

Intel RealSense D435

1.1 Working principle



Figure 1.1: Intel RealSense D435 Camera

Intel Realsense stereo vision technology, differently from traditional 2D cameras, consists in one left imager and one right imager pointing in the same direction and an infrared projector. The distance between the left and right cameras is called baseline. After left and right imagers captured the scene the vision processor tries to find correlation among points between the two images to get the respective depth value in the depth image. Depth is calculated considering the shift of same pixel in the two different images. This value is called disparity and is measured in pixels unit. In order to find homologous pixels between two images, local features inside them are matched up. Near to the camera objects results in having a larger disparity and far objects results in having a smaller disparity. Thanks to triangulation is possible to relate disparity to depth measure for every point in the image. Depth is then calculated from this equations [1]:

$$Depth(mm) = \frac{focal\ length\ (pixels) \cdot Baseline\ (mm)}{Disparity\ (pixels)} \quad (1.1)$$

$$\text{where focal length (pixels)} = \frac{X_{res}(\text{pixels})}{2 \cdot \tan\left(\frac{HFOV}{2}\right)} \quad (1.2)$$

with X_{res} = the horizontal resolution chosen (best one is 848 [2]); Horizontal field of view of left imager (HFOV) = 91.2 °; Baseline = 50 mm.

The camera can provide up to 1280 x 720 active stereo depth resolution and up to 1920 x 1080 RGB resolution [3]. The purpose of the infrared projector is to help create depth maps in textureless scenes projecting a light pattern in all the images. This projector doesn't interfere and so doesn't influence the color channel image.

Since the RGB cameras and the stereo cameras don't provide aligned images, to process RGB and Depth images referring to the same pixel coordinate system makes necessary to align them through a post processing procedure. Without this alignment a pixel at a certain coordinate in the RGB image wouldn't be described in depth by a pixel at same coordinates in the depth image. By doing this procedure the usable areas for the images reduces to 640 x 480.

This camera can be adopted also for outdoors measurement in sunlight scenes. In particular sunlight reduces the sensor noise and tends to bring out textures in objects. Some problems could arise with lens glasses or when image points at the sun. The presence of glares in images caused by sun or light sources reflection on objects could cause holes in depth images maps.

Intel Realsense camera provides a color image, a depth image, a point cloud file and eventually left and right stereo camera images. Images coming from stereo imagers are mono channel images and include infrared projector points. Infrared pattern is not present in the color image making possible to use this image in combination with the depth map for further image processing algorithms. Starting from the depth map is possible directly to get depth of singular pixels since there exists a direct correspondence between color and length.

1.2 Main configuration parameters

Intel provides a RealSense viewer in order to test and set camera parameters. This environment provides also a calibration tool useful in order to calibrate the camera reducing the depth estimation error. Main settings used in the following analysis will be now introduced.

Presets

Presets are collections of parameters configurations devoted to cover all possible different camera applications. Presets can be easily imported or exported during camera usage through json files. Depending on how much

reliability or completeness it's needed in the depth image a certain preset can be chosen. Intel provides some preset briefly illustrated in Table 1.1 [4].

Preset Name	Main usage recommended	Resolution based
High Density	Higher Fill factor, sees more objects. (Ex. BGS and 3D Enhanced Photography, Object recognition)	Yes
Medium Density	Balance between Fill factor and Accuracy	Yes
High Accuracy	High confidence threshold value of depth, lower Fill factor. (Ex. Object Scanning, Collision Avoidance, Robots)	Yes
Hand	Good for Hand Tracking, Gesture recognition, good edges	No
Default	Best Visual appeal, Clean edges, Reduced PointCloud Spraying	No
Edge Map	Extract and map in depth only RGB image edges	Yes

Table 1.1: Main presets specifics [4]

Stereo Module

Resolution

Different Resolution - Frame per Second (FPS) combinations can be used. In general the higher the FPS desired the lower are the usable resolutions and vice versa since enlarging the acquired image dimensions requires more computation time. It's important to underline that the optimal resolution in order to get the best depth image in the D435 camera is 848 x 480 [2]. Lower resolutions can be used but this will degrade the depth precision since matching among left and right images depends on how defined are details in the two images.

Autoexposure

This feature is very important since exposure is responsible for having the best depth image in terms of quality and completeness. Bad exposure, in fact, could generate low density mapping in depth image. Both auto exposure for the RGB camera and for the Depth camera can be set independently. The auto exposure set point is selectable through the "Mean intensity set-point parameter". This parameter decides which luminosity have to have

pixels in a certain ROI in mean. The ROI can be changed manually to get the right illumination for the subject in the image depending on it's position.

Laser Power

Intel RealSense camera uses an "Active" stereo vision because it relies on the addition of an optical projector that overlays the observed scene with a semi-random texture that facilitates finding correspondences. This feature helps in particular in the case of texture-less surfaces like indoor dimly lit white walls. Laser is an additive information used to map better depth especially with far or textureless plain objects. It increases the texture of objects thanks to the random texture projected pattern and, in this way, allows to increase the number of matches between left and right imagers. The distance at which the laser pattern is visible by the imagers depends, of course, on the laser power intensity. For depth mapping of near subjects the laser intensity can be kept low, vice versa to have better characterization in depth of far subjects the laser power intensity has to be increased. Increasing the laser power increases of course energy consumption of the camera but also increases definition of the depth image for far features.

The projector lies between the left and right stereo imagers and is able to turn on always or only when required. It can also be deactivated but this may decrease the depth map quality.

Gain

Gain represents sensibility of the sensor to light. The higher this value the more sensible will be the sensor to light so it will produce a brighter image keeping the same exposure time. This comes not for free since increasing sensor sensibility increases also electronic noise producing a more disturbed scene image and a final degraded depth image.

Advanced Parameters

RealSense maps uncertainties in depth with black pixel values. These are "holes" in the depth map that represent depth data that don't meet the confidence metric and the camera, instead of providing a wrong value, provides a zero value. Major uncertainties are near the object's edges.

Since depth mapping is based on finding correspondences among two images, it's necessary to select some parameters in order to decide which are good matches and which not. This property could be set through pre-sets but for our application can be slightly adapted in order to get an even lower number of not mapped depth pixels. A fundamental parameter is the "DS Second Peak Threshold" that is responsible for limiting matches among images and is related to scores given at best matches and second

best matches. This parameter establishes how much the scores given automatically at the two matches have to be different in order to say that a match is good. Increasing the value of this threshold more depth results will appear as black, vice versa a low value could include wrong matches among repetitive structures like, for example, fences [2].

Depth Units

It's the number of meters represented by a unit of length. By default the camera provides 16bit depth values with a depth unit of $1000\mu\text{m}$ (1mm). This means the max range will be of about 65m. Changing this to $100\mu\text{m}$, for example, allows to report depth to a max value of 6,5 m. This could be helpful in case of near range camera usage because at each $100\mu\text{m}$ is associated a different color in depth map leading to a very high depth resolution. There are different color maps applicable to the depth map: jet, classic, white to black, black to white, bio, cold and warm. Depending on the depth color map desired one of them can be chosen. In conclusion the minor the value of this parameter the major precision will have near objects, at expense of having a smaller overall range of depth representable in depth image.

Depth Clamp Max and Depth Clamp Min

These two parameters allow to restrict the depth represented range to the one specified by these parameters putting dark values in the other.

Histogram Equalization

Histogram equalization makes the colors in depth images change proportionally within the available valid depth pixel range. For example, if the depth image contains valid depth pixels values ranging from 20cm to 30cm, the color will spread proportionally within this range granting best color-depth definition obtainable. The same effect can be obtained activating the Dynamic Visual Preset mode. Disabling this feature will cause the color representation to change based on the minimum/maximum mapping range imposed manually.

Disparity Shift

Through this parameter is possible to decrease the minimum operative distance of the camera but with the drawback of reducing a lot the maximum operative distance. What is before the minimum and over the maximum operative distance is mapped with black values [2].

A-factor

Through this value is possible to compensate errors due to non linearities in the depth calculation procedure. Some optimum of this value are provided depending on the preset used for depth image mapping [5].

Post-processing filters

These filters can help improving depth or RGB images quality but introduce additive computational overload for the camera. The post process time have to be added at the exposure time to obtain the total time needed to have final depth and RGB images. Post processing filters are:

- Sub-sampling

This procedure consists in reducing the resolution of collected images to make them lighter and speed up stereo algorithms. To better preserve details in the original image instead of applying a brutal sub sampling filter, Intel provides "non-zero median" or "non-zero mean" sub sampling techniques. This filters keeps for each pixel of the sub sampled image the median/mean of non zero values of neighbors of that pixel in the old image.

- Temporal Filtering

This filter is used both for closing image black holes and to enforce depth estimates along time. Its main purpose is to counteract temporal noise that consists in oscillations of pixels values around their true value and it works by basing current pixel estimation also into the value of this pixel in past depth images. Temporal noise can be caused by several sources: changes in environment lightning, sensor noise, motion or projector noise. Inside this filter is present also a persistence filter: for every depth missing pixel, the filter will look at the depth images history and will retain the last valid value seen.

- Edge-preserving filtering

This filter is useful to smooth depth noise, making surfaces flatter but preserving edges. If the filter is too much aggressive lots of features will be removed leading to have unusable images. Depth map is scanned in X-axis and Y-axis and back again, twice, while calculating the one-dimensional exponential moving average (EMA) using a parameter that determines the amount of smoothing. This procedure could be speed up by activating the conversion of the image in the disparity domain where disparity is $1/\text{distance}$. This allows also to use the same algorithm parameters for all the depth image compensating the fact that in stereo cameras the depth noise increases as the square of the

distance. By not applying this transformation we would obtain over-smoothing for near-range data, while under-smoothing for far range data.

- Holes filling

Depth holes may result from:

- A Occlusions: the left and the right images do not see the same object due to shadowing;
- B Lack of texture: in case of texture-less surfaces the matching could result very difficult. Here infrared light pattern could help in creating the depth map;
- C Multiple matches: when there are multiple equally good matches;
- D No signal: this happens if the images are under-exposed or over-exposed;
- E Below minimum z distance: this happens when objects are too near the camera. In this cases something can be done reducing the baseline as described accurately in Disparity shift Paragraph.

In some cases it's necessary to eliminate all holes in depth image and this can be done in only one image scan. Three methods for filling are implemented by filling the hole pixel with:

1. Left valid pixel value;
2. The biggest (farthest away) among the valid five upper left and down pixel values;
3. The smallest among the valid five upper left and down pixel values.

Overall filters timings can be resumed in the tables below [6].

Sub-sample factor	Sub-sample Time	Depth to Disparity Time	Spatial Filtering Time, ms (+ hole filling*)
1	0	4.09	22.18 (+ 4.38)
2	4.84	1.83	8.00 (+ 1.31)
3	5.81	0.95	3.92 (+ 1.11)
4	3.02	0.62	2.61 (+ 0.56)
5	2.91	0.44	2.01 (+ 0.54)
6	2.85	0.34	1.34 (+ 0.39)

Temporal Filtering Time, ms	Disparity to Depth Time, ms	Hole Filling Time, ms	Total Processing Time, ms (+ Spatial hole filling)
3.13	3.49	3.09	36.96 (+ 4.38)
1.32	1.41	1.17	19.24 (+ 1.31)
0.76	0.83	0.68	13.22 (+ 1.11)
0.53	0.57	0.47	7.97 (+ 0.56)
0.39	0.40	0.33	6.60 (+ 0.54)
0.30	0.30	0.26	5.47 (+ 0.39)

Table 1.2: For 1280 x 720 image on i7 - 6950X processor; * Hole Radius equal to 2

RGB Module

Color imager includes different settings for image acquisition: brightness, contrast, gain, hue, saturation, gamma correction and sharpness. It's provided of an auto exposure and an automatic white balance system. Also a sub-sampling post-processing filter is present.

On chip calibration tool

This tool is used to correct intrinsic (lens shift) and extrinsic distortions (bending or torsion of the support on which the sensors are mounted). Both distortions cannot be corrected simultaneously but only one at a time. It's possible to set different test times based on the type of correction needed. It's important to keep into account that the sensor will remain busy during this time period in order to perform calibration.

0 = Very fast (0.66 seconds), for small depth degradation

1 = Fast (1.33 seconds), for medium depth degradation (recommended setting)

2 = Medium (2.84 seconds), for medium depth degradation

3 = Slow (2.84 seconds), for significant depth degradation

As soon as the calibration is finished the sensor uses the new estimated parameters. It is necessary to check if the new calibration has actually brought greater precision before saving the new parameters permanently on the device. In general, a correct calibration should reduce the "RMS Sub-pixel Error". This error is based on disparity so it's independent from the distance and not affected by the fact that the distance measurement noise

increases with distance. The Subpixel RMS Error requires to be calculated pointing the camera in front of a wall and is calculated as:

$$RMS = \sqrt{\frac{\sum (D_i - D_{pi})^2}{n}} \quad (1.3)$$

$$\text{where } D_i = \frac{BL \cdot FL}{Z_i}; \quad D_{pi} = \frac{BL \cdot FL}{Z_{pi}}; \quad (1.4)$$

with

Z_i - depth range of i-th pixel (mm)

Z_{pi} - depth of Z_i projection onto plane fit (mm)

BL - optical baseline (mm)

FL - focal length, as a multiple of pixel width

Calibration can be carried out with a generic image, therefore it's not necessary to have a particular surface on which to apply calibration (like a white wall or a certain pixel pattern). This tool also provides a "Health-Check" indicator. Based on the size of this value it's possible to evaluate how much the actual calibration deviates from the best one.

Tare calibration tool

The purpose of this tool is to set the ground truth reference distance for the camera. In general the procedure is to place the camera in front of a flat surface and measure the distance from the left camera to the surface. Once this measurement has been entered the calibration is completed. In this case it's important to set it with "High Accuracy Depth" in the preset mode [1].

Chapter 2

Basler blaze 101

2.1 Working Principle



Figure 2.1: Basler blaze 101 Camera

A simplified explanation of the camera working principle is now introduced. The main idea behind Time of Flight technology is to count time that packets of light sent takes to come back to the camera. Light source is switched on and off alternatively by the control unit while the shutter opens and closes two times. Shutter opens as soon as the light source is set to on and closes when it's set to off, creating image S_0 . After the light source is set to off the shutter reopens and recloses when the light source is reset to on, creating image S_1 . In S_0 near objects will result very lightened since more the light is reflected to the camera more the subject is near to the camera. Vice versa in S_1 closer surfaces result darker and far ones result brighter since no light is sent to objects [7]. Distance can be then measured by the following equation:

$$d = \frac{c}{2} \cdot t_p \cdot \frac{S_1}{(S_1 + S_0)} \quad (2.1)$$

where

t_p is the duration of the light pulse

c is the speed of light

S_0 and S_1 are the collected light in the two different time spans

If objects are very near to the camera, no light is captured in the second shutter period then $S_1 = 0$ and the equation gives $d = 0$. Instead, if no light is captured in the first shutter period, objects are very far and equation yields:

$$d = \frac{c}{2} \cdot t_p \quad (2.2)$$

This distance consists in the largest distance measurable and shows that with ToF technology the maximum measurable distance is strictly dependent on the light pulse duration. For example, if t_p lasts 47 nanoseconds, then up to 7 meters can be measured.

The ToF working principle relies on time of travel of projected light pulses from transmission to return to the sensor. Ambient light doesn't influence the performance of this camera. An optical band-pass filter behind the lens allows only wavelengths generated by the camera's light source to pass through. This guards the sensor against overexposure by a disturbing extraneous light. The near infrared light used by this camera could produce intensity images greatly different from what can be seen with visible light. Since light transmission is influenced by all the detours and shortcuts the light takes along its path, depth estimates are not impossible to error.

Basler camera provides one depth image, in Mono 16 bit or 8 bit color format, a Mono 16 bit confidence map, a Mono 8 bit intensity map and a point cloud file. Intensity images are simply environment brightness images collected thanks to laser illumination. In depth images color maps are applied in the range specified by the user, the smaller the range the higher depth definition in terms of color gradings. This depth range is fully customizable depending on the application. Red surfaces are close, blue surfaces far. Confidence map provides a grayscale image in which pixel brightness is proportional to estimate trustability in the correspondent depth image pixel. Here alignment between depth image and intensity image is not required, images provided are already pixel aligned. Also in this camera the color-distance relation exists and can be exploited to find each pixel distance.

2.2 Main configuration parameters

Exposure of the camera can be set between 100 and 1000 μs in steps of 10 units. This camera can work even in dark ambients since exploits laser light both to illuminate the scene and to map depth. Confidence image map can be manually thresholded in order to set to zero the darkest (untrustable) estimation values.

Basler camera provides high freedom in choosing parameter combinations. Both exposure and depth range mapping can be set independently and with high precision. In Helios 2, that is another time of flight camera produced from a Basler competitor, depth ranges can be chosen from 6 fixed ranges intervals. Exposure instead can be set only to 3 values: 62.5 μs , 250 μs or 1000 μs [8]. Basler in terms of settings gives better freedom respect to Helios 2 camera since having the complete possibility to establish mapping depth range allows to have the desired color mapping for the depth image. Reducing or increasing the range allows respectively to obtain more or less precision from the colormap. The same can be said for exposure time: Basler gives higher flexibility in setting the exposure time, provides better well illuminated scenes and, as a consequence, better estimates for the depth color map.

An auto exposure algorithm is not provided for this camera, this implies that each time application changes or usage distance changes, the exposure has to be manually set. Anyway in our application once good exposure in a certain configuration is found then, even if ambient light changes, it's not necessary to readapt exposure since filters on camera lenses make it unsensible to these frequencies. A good exposure is reached when the Filling Rate of the depth image is the maximum possible and values in this image have little oscillations. Alternatively the confidence map can provide useful information to evaluate the ROI best illumination: the brighter the image the more confident are depth measurements.

$$\text{Filling Rate} = \frac{\text{Mapped depth pixels}}{\text{Overall number of pixels}} \quad (2.3)$$

Intensity Calculation Method

The intensity calculation method establish the synchronization between depth and intensity images. Two methods are available:

- Method 1: This is the default setting. In scenes containing motion may be notice a lag between depth image and intensity image.
- Method 2: With this setting, the depth image and the intensity image are in synch but the resulting intensity image will be darker and contain more noise.

Fast mode

The Fast Mode feature allows to achieve the maximum frame rate of 30 fps. However, enabling it also reduces the measurement accuracy. Normally, the Basler Blaze camera takes eight subframes per image at different frequencies. If the Fast Mode is enabled, only four subframes are taken. If the Operating Mode parameter is set to Short Range, the subframes are taken at high frequencies (100 MHz). If the Operating Mode parameter is set to Long Range, the subframes are taken at low frequencies (15 MHz).

Fast Mode also reduces motion artifacts that are caused by the eight exposures used in normal operating mode. The downside is that the resulting depth data is going to be less accurate in particular when working in Long Range mode [9].

Outlier Removal

The Outlier Removal allows to remove pixels that differ significantly from their local environment. If a pixel's depth differs too much from the depth of its neighboring pixels it's called outlier and will be classified as invalid. The intensity image is not affected by this feature. Points removed assume, in the depth map and point cloud, special values decided by the user. In the Confidence image, outlier pixel values are set to 0 [9].

Operation modes

Two operation modes can be chosen: Short Range and Long Range. Using the Long Range mode sets the exposure time at its maximum value 1ms, viceversa in Short Range the exposure time is set to its minimum that is 100 μ s. Anyway exposure time can be changed manually through the Exposure Time parameter. Automatically these setting select also the maximum and minimum Working Ranges. To adjust the effective working range the Depth Min and Depth Max parameters can be used. Each camera frame is composed of eight subframes. Four of these subframes are taken in Short Range (with a frequency of 100 MHz) and the other four subframes are taken in Long Range (with a frequency of 15 MHz). This combines the higher accuracy of the Short Range mode with the longer range of the Long Range mode.

In Table 2.1 is represented the precision given by these two parameters related to their maximum and minimum Working Ranges [9].

Also in this camera are present Temporal and Spatial filters and a gamma filter to betterize depth images. A thermal drift correction functionality is included in the acquisition software and in the camera libraries.

	Short Range	Long Range
Range	0.3 to 1.5 m	0.3 to 10 m
Accuracy	± 5 mm (0.5 to 5.5 m)	± 5 mm (0.5 to 5.5 m)
Temporal noise	< 2 mm (up to 1.5 m)	< 5 mm (up to 5.5 m)

Table 2.1: Camera Precision in Long and Short Ranges

Gamma Correction

Enlighten the darker part of the intensity image if required with a gamma correction with gamma equals to 0.4 .

The result value is scaled to preserve the original full-scale 16-bit image.

$$OUT = (2^{16} - 1) * \left(\frac{IN^{0.4}}{2^{16}} \right) \quad (2.4)$$

where OUT = result; IN = 16 bit input pixel intensity

Temporal Filter

The filter happens recursively adding the old depth image values weighted by $k/255$ to the new value weighted by $(k - 1 / 255)$. "k" is the Strength parameter that can be customized depending on how much the user wants the current frame to be influenced by the old one. This filter is able also to detect motion searching for high changes of pixel values. In that case the corresponding pixels are reset and reinitialized to avoid motion artifacts [9].

The following equation shows how this filter works:

$$OUT = \left(\frac{k}{255} \right) * STATE + \left(\frac{1 - k}{255} \right) * IN \quad (2.5)$$

where OUT = result ; IN = new frame; STATE = old frame; k = filter strength (value of the Strength parameter)

Basler camera is influenced by different light undesired behaviours and sensor intrinsic issues that could compromise or make imprecise depth measurements.

Multiple Reflections

Measurements need that light has to be reflected only once by environment surfaces. If it's reflected multiple times the time of flight of this packet of light increases and consequently generates a more far measurement respect to real one. This is the case of concave objects or corners in which depth estimation is distorted and wrong with respect to real one. Reflective surfaces may deflect away all the light beams leading to completely wrong depth estimates. Sometimes could happen that these kinds of reflective surfaces deflect beams directly inside the camera causing a phenomenon of over exposure that leads to not estimated depth areas in depth image. Diffusely reflective objects are so preferable for this camera usage.

Scattered Light

This phenomenon happens when unwanted reflections within the lens or behind it happens. Surfaces that aren't even in the sensor field of view may scatter light into the lens, this phenomenon generates washed out intensity images with low contrast.

Working Area

Temporal arrangements of light pulses and electronic shutter are designed for specific distances ranges. Surfaces that are not inside these ranges show not measured or incorrect values in depth image. Instead the intensity image could be out of focus or poorly illuminated.

The intensity of light reflected is reduced by the square of the distance so near objects result bright while far ones are dark. A good exposure parameter setting is required in order to get complete bright and dark information. The intensity of the light source must be sufficient to illuminate well subjects even at large distances otherwise high noise, oscillations and imprecisions will occur for depth estimates. Depth of focus of the lenses may generate indistinct borders between foreground and background. In depth map these areas could be unmapped.

Solid Angle

In the marginal regions of the image light source intensities are reduced. Objects and surfaces in these areas could result underexposed and less precise.

Ambient Light

An optical band-pass filter only allows the camera light source to pass through the sensor and makes the camera insensible to artificial light that contains small components passing this filter. Daylight, instead, could cause problems since it has components in all the light spectrum. This light could increase light collected by the sensor and generate bad or highly disturbed frames.

Degree of Reflection and Transparency

Depending on the reflectivity of the environments the optimal exposure time has to be chosen. In case for example of both high reflective and not reflective subjects a fine exposure time is essential to get a good image. In this case a possible alternative could be to shoot two times the same image with different exposures.

Temperature

Camera is also sensitive to temperature since if it's high and not stable, noise and imprecision will occur in measurements. Temperature in fact affects electronic and timings inside the camera chip [10].

Main specifics of the cameras are resumed in Table 2.2.

	Basler blaze 101	Intel RealSense D435
Measuring Method	Time of flight	Stereo matching
Resolution	640 x 480	up to 1280 x 720 Depth and 1920 x 1080 RGB
Working range	0.3 m - 10 m	0.17 m - over 10 m
Accuracy	5 mm (0.5 - 5.5 m)	$\leq 2\%$ (up to 2 m)
Frame Rate	30 fps	up to 90 fps
Mono/Color	Mono	Color
Temporal Noise	$< 1\%$ (up to 1.5 m)	$\leq 1\%$ (up to 2 m)

Table 2.2: Main cameras specifics [3, 11]

Chapter 3

Application target and specifications

Main focus of this work deals with pharmaceutical and medical products. Some of these products are difficult to be handled with mechanical orientation procedures and it's necessary to introduce robot picking techniques to place these products in their boxes, ready to be closed. Through the help of computer vision, robot arms need to know objects positioning, orientation and direction in space in order to catch them when they are in their working area. In particular we are talking about objects like syringes, pipette, blisters, small bottles and small jars. Products have different shapes, colors, they can be textured or not textured, lucid, opaque, transparent, flowpacked or not. Actually machines processing these products are equipped with 2D cameras able to locate and orient objects passing on conveyor belts. What is required is to put in evidence possible advantages that commercial 3D cameras could carry. Principal objectives of the algorithm designed is to grant soft real time results, this means that they have to be computationally enough fast to provide products positions to the robot before they pass beyond the robot working area. Otherwise robot won't be able to catch them since they are too far from it. There are no requirements about quality check of objects involving for example shapes anomalies or dimensions measurements.

In all this work, objects that can be kept but are discarded will be called false negatives and object that cannot be kept but are anyway caught are called false positives. In general a strong required constraint is to pick up only free objects, no positions of occluded objects have to be sent to the robot since it couldn't be able to catch them. Those objects, that are not sure to be free with certainty, have to be discarded. Following this considerations false negatives are accepted but false positives have not to be present.

It's not essential to pick up all the objects on the conveyor belt since objects that are not grabbed are recollected and moved back at the start of the conveyor belt. In this way no products are effectively lost if not picked up by the robot arm, only the efficiency drops. The more objects are detected as free and picked up the more the efficiency of the algorithm designed increases. The objective of the algorithms is both to detect as free objects that are isolated from the other but also to detect free objects that are on some other ones. The overlapping management is the main and the major challenge that have to be managed in this application.

It's important to take into account among the methods designed their flexibility and adaptability to different and various objects. Ad hoc solutions have to be avoided since the creation of one different algorithm for the localization of each object would require a lot of time. Instead adaptable methods requiring small parameters changes from one object to another could speed up development processes.

Chapter 4

Products qualitative analysis

In this chapter are done some preliminary analysis on cameras performances on products. First of all both Basler blaze and Intel RealSense cameras images features and characteristics are explained to make following images more understandable. Examples of well depth mapped objects are used to show how images are expected to be. Then cases of objects giving mapping problems are introduced, highlighting main problems that can arise using these cameras. This is fundamental for future object analysis to be aware and able to justify and mitigate possible unwanted distortion effects. In this way it's possible to check weaknesses of cameras and eventually, in case the depth map doesn't carry significant or well defined information, exclude products from further considerations. Since the aim of this work is to enhance cameras advantages, it could be helpful to initially discard these objects for which 3D cameras doesn't make the difference with respect of 2D ones.

These products are already processed through 2D images algorithms so it's not required to provide an algorithm dealing with all objects. The focus of this chapter will be, then, on enhancing cameras wrong or unexpected behaviours.

4.1 Basler blaze 101



Figure 4.1: Intensity Image

Intensity image, in Figure 4.1, corresponds simply to the brightness grayscale image of the environment illuminated by laser light.

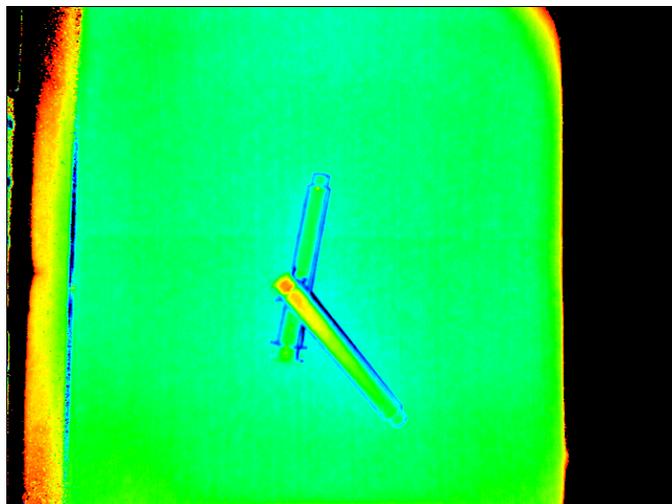


Figure 4.2: Depth Image

Depth range map corresponds to a color mapping of the depth measured for each pixel (Figure 4.2). It exists, then, a Basler function that translates the depth range into the color range. By measuring the time of flight of light, the camera calculates a depth value in each pixel. This depth is consequently remapped in a color scale going from blue to red (where red

is near and blue is far), with black pixels unmapped points. All depths between these two colors are mapped with RGB intermediate values. The color map is automatically rescaled inside the value specified by the user in Basler Viewer or by Basler libraries. Out of this range pixels are mapped in black. This image is available also in 18 bit grayscale format where dark pixels are near and white are far.



Figure 4.3: Confidence Image

About the confidence image, in Figure 4.3, a depth trustability value from black to white is assigned at each pixel in a 16 bit gray level scale. Values represent confidence of estimation in the correspondent depth pixel values: the darker the values the less confident it's the estimation. For example, black objects in the intensity image are badly estimated while white ones are very well estimated since they reflect better light back to the camera. This features will be used in algorithm described in next paragraphs.

4.1.1 Transparent objects

Major part of objects analyzed is made of lucid or transparent plastic. Some pipettes, for example, are made completely of transparent materials, the same happens with external parts of syringes that are made of transparent lucid plastic. Transparent plastic objects result almost invisible to the cam, as it's possible to see from Figure 4.5.



Figure 4.4: Transparent objects Intensity Image

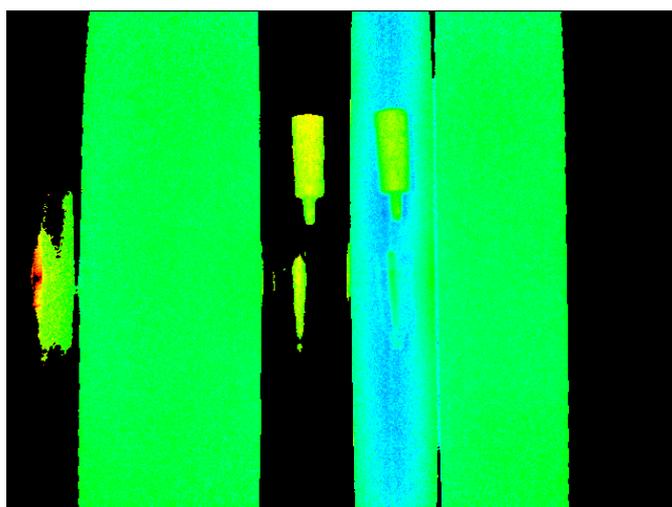


Figure 4.5: Transparent objects Depth Image

In the high part of the intensity image (Figure 4.4) there are two white plain pipettes clearly visible. Instead, in the lower part, there are two of them transparent. These pipettes have almost the same shape, the only things that change are their color and degree of transparency. The two transparent objects are quite visible from the intensity image but anyway not so visible as the white homologous one. White objects could be segmented easily only using the intensity image and applying an edge detection procedure, exploiting the fact that dark background enhances contrast. Vice versa transparent objects would be difficult to be segmented since transparent plastic creates low contrast between background and object color. So

it's possible to detect them by the intensity (brightness) image but not as easy as the white not transparent ones.

Looking at the depth image in Figure 4.5, a different and apparently strange result is obtained for transparent objects that results almost invisible if only looking at the depth map.

Before trying to explain this behaviour, one consideration about don't mapped black conveyor part has to be done. Basler camera provides the possibility to threshold the confidence image in order to eliminate from depth map objects that are mapped with high uncertainty. By setting the threshold at about 6000 (so at a quite low value), black parts in the intensity image stop to be mapped in the depth one. An explanation of this fact is that black materials, especially if opaque, absorbs all light projected on them. This is something not expected by the camera since it bases its depth estimation on the time in which the light is reflected back to itself. Black points absorbing all light would result then difficult to be mapped or would be mapped wrongly. Without using the threshold, Basler camera tries to map in depth black conveyors doing quite near to true estimates (Figures 4.6, 4.7).

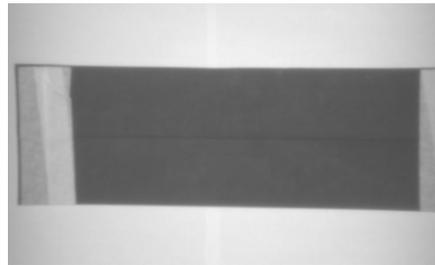


Figure 4.6: Black conveyor belt Intensity Image

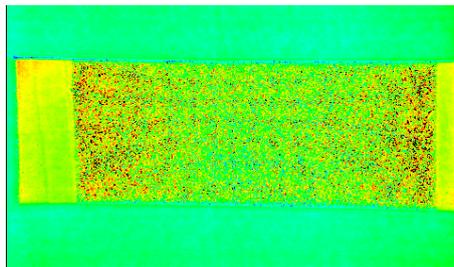


Figure 4.7: Black conveyor belt Depth Image

Anyway these estimates are not so confident, they are very unstable and oscillating due to fact light reflected by black material is almost fully absorbed by itself. The fact that these estimates are not so confident is double checked from the fact that even with low confidence thresholds these

estimates are no more mapped in depth. This explains why white conveyor part is mapped in depth and the dark conveyor part no. Black objects are then very difficult to be well mapped in depth by this camera and could cause problems, for example, with the depth map generation of objects containing black textures or black components.

One example of an object very difficult to be treated among the ones examined is a pipette with a black head and white plastic body. From its depth images the black head results badly mapped and the body instead clearly visible. A solution found for this kind of black objects is to increase the camera exposure parameter. In this way dark objects are better illuminated increasing the quantity of light reflected. Estimates result more accurate and black parts in intensity images are well illuminated. On the other hand overexposure of white objects can arise and white parts couldn't be mapped completely in depth due to this fact. Further and detailed information about saturation phenomenon will be given in Chapter 6.

This black not well mapped behaviour can be seen not only as a problem but also as an advantage since white objects are very well mapped with the black background and a first objects segmentation is provided for free. In fact, Basler algorithms explained ahead make a positively use of this behaviour.

Looking back at the depth map in Figure 4.5, while white objects are well mapped, transparent ones result almost completely unmapped in both white or black conveyor. It's evident that Basler camera is unable to detect transparent objects even if in the intensity image these objects are visible.

This behaviour is not connected to conveyor color since the image contains the same objects placed on a dark and a white conveyor of the same type. Current conveyor is reflective and lucid but tests have been done also with plain dark opaque conveyors. Also in these cases transparent objects resulted not mapped in depth even if visible in the intensity image.

This depth invisibility behaviour can be attributed to the fact that projected laser cross the objects without being influenced or reflected by them. This is confirmed by the fact that white objects of the same type are very well mapped with respect to transparent ones for which no trace is present in depth image. It's possible to conclude that transparent materials result invisible to this camera. In general this camera doesn't carry additive information for this kind of objects with respect to a 2D camera. Objects presenting this transparent characteristic won't be considered in further analysis.

As introduced in Chapter 2, the intensity image is based on near infrared light illumination and then the image collected by the camera can be different from what seen by humans in the visible spectrum. In this case, there are not too many differences from the image collected by the Basler camera and a black and white image of the same scene collected from a 2D camera. Instead, with certain kind of textured objects, like in Figure 4.9, by

filtering out with the optical pass-band filter only near infrared wavelengths, it can happen that textured part disappears. In Figure 4.8 and 4.9 the same image is kept respectively with Realsense RGB camera and with Basler one, showing that it's not always told that textures are visible using the Basler camera. Knowing that sometimes a camera couldn't detect textures will be fundamental for the future identification algorithm that has to be designed to work for both cameras, independently from objects textures.

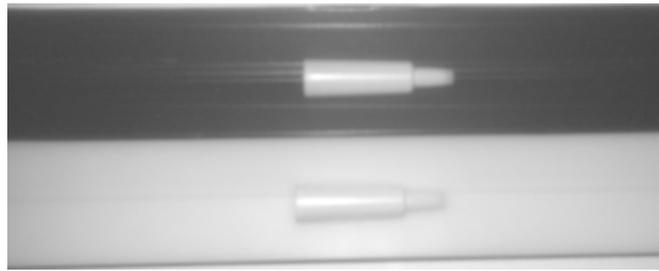


Figure 4.8: Basler blaze 101 - textured objects Image



Figure 4.9: Intel RealSense - textured objects Image

4.1.2 Flowpack objects

The phenomenon of having transparent objects not mapped in depth is not always true. Considering for example objects inside flowpack (Figure 4.10), plastic folds generate local reflections of rays back into the camera. Plastic in this case behaves locally like a mirror reflecting rays back to the camera. Objects inside flowpack are mapped well in depth but plastic could also cause local reflections to be mapped in depth. This is something that in the previous case didn't happen and is linked to the incidence angle of light projected by the camera.

By looking at the final depth map (Figure 4.11), it's impossible to tell which point belongs to the object and which is only an unwanted depth estimation coming from transparent plastic. Here black conveyor is fundamental since it doesn't reflect light back to the camera. A white conveyor would generate worse results since it would increase the light reflections. In that

case plastic reflections could increase, creating more depth artifacts in the depth image. We would also lose the advantage of having a light absorbing background that filters out for free depth information of the conveyor.



Figure 4.10: Flowpack Intensity Image



Figure 4.11: Flowpack Depth Image

Anyway, even using black conveyor depth map results confusing. It cannot be told which parts of the depth map belong to the object and which don't. No direct object segmentation can be done from depth, the maximum help that depth map can give, in this case, is to filter out from intensity image the areas in which objects are. In this way a detailed search for the object can be done only in areas mapped in depth. As can be seen from Figure

4.10, even filtering out parts around the object, an object search procedure results difficult since plastic doesn't clearly show the object. In conclusion, the Basler camera doesn't grant so many advantages with flopacked objects with respect to 2D cameras. In addition, using 2D cameras in combination with polarized filters for flowpack objects allow to filter out transparent plastic material. In this way only what is inside the pack can be seen without the disturbance effect of plastic around. Instead, the usage of a polarized filter with Basler could compromise depth mapping quality.

4.1.3 Local saturations

Lucid objects, like plastic syringes, present high reflection capacity. If operating at short distances, laser beam power is very high especially in the central part of the image. In some parts of the image local over saturation may arise as it's represented in white areas in Figures 4.12. From the intensity image it's possible to see that local light saturations arise on the center of the objects. It also results difficult to predict areas in which this phenomenon will arise since it's strictly linked to object position in the camera field of view and on the object orientation in space.

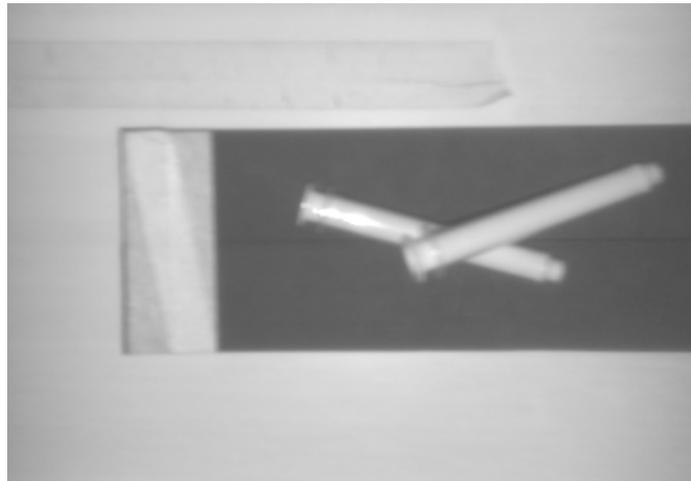


Figure 4.12: Local saturations Intensity Image

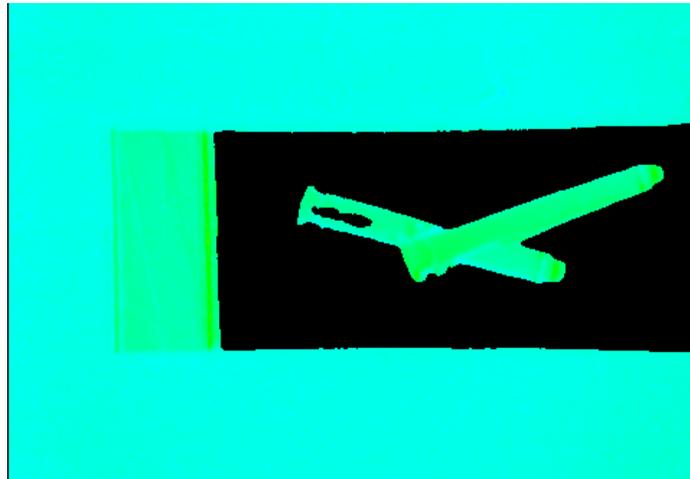


Figure 4.13: Local saturations Depth Image

This phenomenon can be explained due to a large part of the light is directly reflected back to the camera causing saturation areas. Direct reflections are linked to the fact that these objects are not diffusely reflective objects. From the point of view of the quality of the intensity image this is not a big problem, the real trouble arises in depth image where black holes come out in correspondence of intensity saturation areas (Figure 4.13). In other words saturated white areas in the intensity image are mapped with black points in the depth map image.

Nothing can be done in these cases, except choosing very accurately the exposure time: the lower it is the smaller are saturation areas. Lowering too much this value can cause bad overall depth estimates since too dark intensity images lead to having more uncertain and oscillating depth estimates.

A solution of this problem will be proposed in the Results Chapter that involves the attempt to fill these holes to reconstruct in the best way the depth image. These can be done exploiting the fact that the confidence map has zero values in correspondence of these depth mapping holes.

4.1.4 Multiple reflections

As explained in Chapter 2, around objects contours multiple reflections of projected light beams could happen causing wrong depth estimates. Light projected, that is deflected multiple times in corners or concave objects, spends more time flying with respect of going back directly to the camera sensor. Then in correspondence of objects edges, depth estimation will be more far respect to the ground truth depth values, since the higher the time spent the more far is the subject observed. This phenomenon can be seen in depth images (Figure 4.14) but becomes particularly evident looking at the point cloud image (Figure 4.15). In the example shown the object contour

map goes under the ground for the multiple reflection effect just described. The background assumes a green/yellow depth color that changes and tends to red in the central part of the syringe as depth measured reduces. Instead, around the syringe shape, blue color contours are present. As told before blue represent the most far measured depth by the camera, also more far than green. This means that blue parts are deeper respect also to green ones meaning the object contours are sinking in the ground. The correspondent object point cloud image (Figure 4.15) reflects the behaviour just described in all of the contour parts.

Since objects considered are small, this distortion effect compromises evidently the object depth mapping. Only their central parts result useful and reliable for eventual depth considerations.

In case of having for example boxes in a palletizer, their dimensions will be higher with respect to their contour bad mapping effect and so the sinking phenomenon described would be negligible. The solutions implemented for nominal cases, as will be explained later, are able to compensate for this behaviour using black opaque conveyor belts.

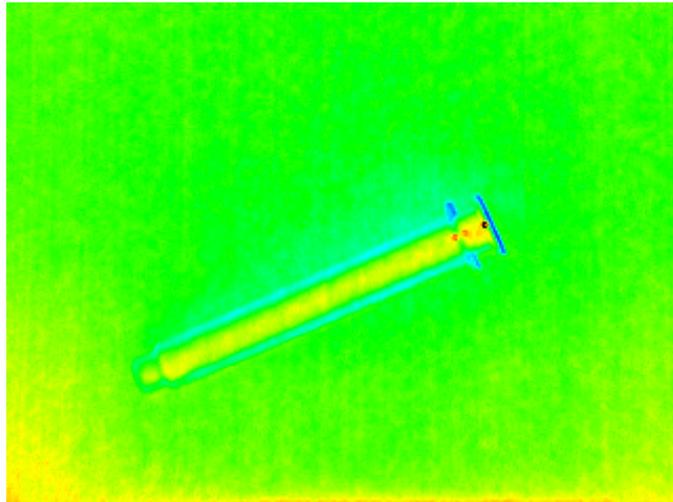


Figure 4.14: Syringe Sinking Point Cloud Image

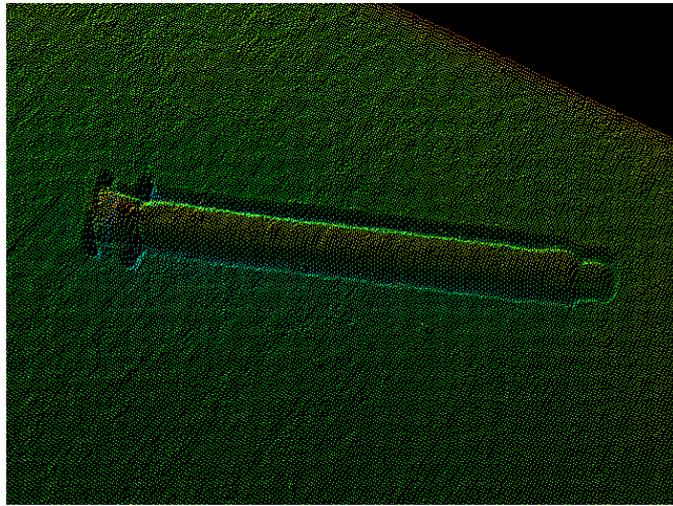


Figure 4.15: Syringe Sinking Depth Image

4.2 Intel RealSense D435

Intel Realsense provides mainly two images, the RGB image and depth image as can be seen in Figures 4.16, 4.17.



Figure 4.16: Intel RealSense D435 Flowpack RGB Image

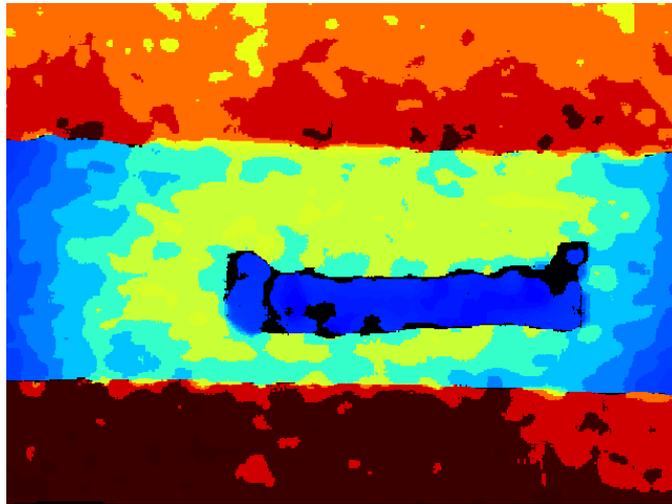


Figure 4.17: Intel RealSense D435 Flowpack Depth Image

Both RGB and depth cameras parameters like exposure, gain... can be set independently in each relative module. The used depth image color mapping principle results inverted respect to Basler camera. Red is far and blue is near; black pixels are unmapped values due to too low confidence in estimates. The color map is not fixed as in Basler camera and can be chosen among different methods (more details in Chapter 1). Depth measures stays equal, the only thing that changes is the color with which a certain distance is represented in depth image. The mapping color range, in case of Histogram Equalization enabled, is automatically clamped to maximum and minimum depth values measured for each camera frame. Otherwise it's possible, like in Basler camera, to set a depth fixed interval in which colormap values are spread. In this case (Figure 4.17) Histogram Equalization mode is disabled.

4.2.1 Flowpack objects

Intel Realsense camera presents similar behaviour of Basler in mapping flowpack. Flowpack objects are almost impossible to be mapped as can be seen in Figure 4.17 . The camera tries to map in depth both the object and the transparent plastic parts. The major problem is that a lot of dark pixels are present meaning that the depth estimation in these points is impossible. It's not clear, then, where effectively is located the object since near contours the amount of estimation uncertainty increases. No segmentation based on the color can be done and the object results difficult to be found through the RGB image since transparent plastic makes difficult to see it.

Now the difficulty in mapping depth points can be attributed to the fact that left and right imagers are differently hit by light reflections on plastic. This lead to have two different left and right images with lot of miss

matching or not very reliable matchings. The resultant depth image is full of depth holes even if the preset used is the High Density one. Using this preset with normal objects causes the image Filling Rate to increase evidently. In this case this doesn't happen since patterns projected in plastic material and strange light reflections generate unmatchable left and right images. In conclusion, depth image is not exploitable or at maximum can be exploited to know where objects are approximately located.

4.2.2 Transparent objects

One different result respect to Basler camera is obtained in case of objects considered in Figure 4.18 and 4.19. These kind of objects are almost not mapped in depth with Basler camera. Instead, with RealSense camera and High Density preset, these objects are mapped well as shown in Figure 4.19. Even if these objects are transparent, the laser pattern is not too distorted by transparency and good matches between left and right stereo images are found. In conclusion, RealSense camera can be used with these kind of transparent objects, it grants reliable depth frames that are compatible with algorithms that will be described in the next Chapter. This result is very important since a transparent object can be mapped in depth by at least one of the two presented 3D cameras. This underlines also the fact that the two technologies have different behaviour with materials, strictly linked to the technology used.

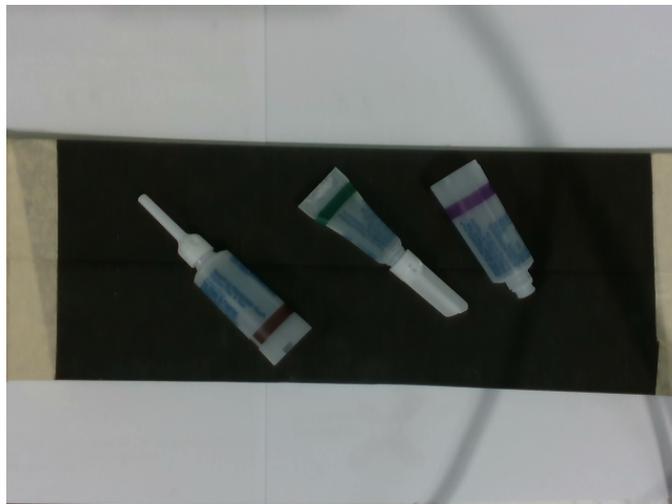


Figure 4.18: Transparent objects RGB Image

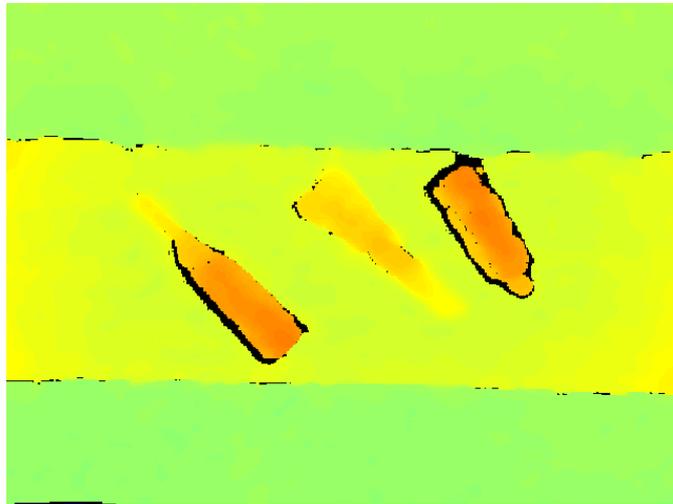


Figure 4.19: Transparent objects Depth Image

Flowpacked objects won't be then treated in further analysis since both cameras fatigue to map them in depth. 2D camera solutions in combination with polarized filters have to be adopted in these cases. Intel RealSense camera is preferable with respect to Basler one since it provides more reliable depth images and with higher Filling Rate.

The Filling Rate will be carried always to 100 % through Holes Filling post processing procedure. This process will be described in detail in Chapter 6. The algorithms designed have to be as immune as possible to Basler sinking contours in depth maps and research of objects have to be independent of eventual objects textures. Also Intel RealSense camera fatigue to map contours in depth, in Chapter 6 will be provided algorithms insensible to these bad depth mappings.

4.3 Nomenclature

Along this work, some terminology is used to be more clear and have common nomenclature for further statistics and analysis:

"Shape matching false positives" happens when an object is wrongly found by the shape matching algorithm in an area in which no objects are present. This behaviour is mainly due to the fact that too low thresholds in the shape matching algorithm are set or misleading environment shapes tricks the matching algorithm along the search. Low thresholds are sometimes chosen when it's necessary to find all objects in a scene and the objects are difficult to be found since, for example, they are hardly occluded and their shapes are not so visible in the mono/RGB collected images. In this

case threshold parameters can be lowered but shape matching false positives may arise. The objective of identifying also the occluded objects is achieved but shape could be matched even in a place in which no object is present. Along this work some techniques will be introduced to counteract this event. These techniques are one of the major advantages given by 3D cameras technologies.

On the other hand it could happen to have objects that have low contrast respect to the background. An example could be white plain objects on a white conveyor belt. In this case even the model extraction could be very difficult since objects contours are not sharp enough. Noisy pixels areas or unwanted shadows can cause mistakes in the research procedure causing objects to be identified where they aren't. For this reason in the Results Chapter a deep analysis about conveyor colors and features adopted will be presented.

The "shape matching false positives" phenomenon has to be avoided since it's not accepted to send to the robot arm objects positioning completely wrong. The positioning has to be given only when there is 100 % certainty to have detected a free object, not wrong matched or not occluded. Anyway "shape matching false positives" have not to be confused with simple "false positives".

"False positive" term is related to an object that an algorithm wrongly classified as grabbable because it's occluded. The object exists and is located exactly where the shape matching algorithm found it. False is related to the classification and not to a wrong positioning. If an occluded object is classified as free, then a false positive arises. This is mainly related to an error caused by wrong design of algorithms and/or wrong parameters settings. False positives have to be always avoided in our analysis since there isn't the possibility to pick up occluded objects. The robot arm, in fact, positions it's vacuum cap end effector in correspondence of the object barycenter but isn't able to catch it since another object is inside of it. In further analysis will be introduced some considerations about setting up parameters to avoid having false positives.

"False negatives" are those objects that are classified as not grabbable by the robot arm but they are, instead, free. Since the requests for the current applications are to avoid at all false positives, the false negative will increase as a consequence. When, for example, it's difficult or ambiguous to establish if an object is free or occluded the choice will always fall on occluded. In this way false positives won't be present but the number of false negatives will increase. One main objective will be to reduce or avoid even false negatives. In the Results Chapter some assessments about best methods for having low false negative rates will be given.

4.3.1 Algorithms efficiency parameters

Once false positives and false negatives definitions have been explained, it's necessary to introduce two parameters that will be used to quantify results in next Chapters. These parameters are called Precision and Recall and are defined by the following equations:

$$\text{Precision} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}} \quad (4.1)$$

$$\text{Recall} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}} \quad (4.2)$$

Both of these parameters can assume values between 0 to 1 and can be easily converted to percentage values (multiplying them by 100). The Precision parameter specifies the quantity of objects that have to be grabbed with respect to the total number of objects grabbed, comprising both real free objects and occluded ones. Since the assumption done in this work is to avoid picking up objects that are not free, the algorithms are designed to avoid having false positives. The precision value, in all the described algorithms will assume 1, meaning that only objects that are for sure free, are grabbed.

The Recall parameter is an index of how much objects are classified as grabbable respect to all the objects that are free and could be picked up. Objects that for different reasons are free but are not detected as free will increase the false negatives number decreasing the Recall value. If the 100% the free objects are marked as grabbable then the Recall parameters will assume the value of 1. As soon as some false negative is found, the Recall parameter decreases.

These parameters are useful to quantify effectively how reliable and efficient the algorithm designed are. By these two indexes, then, it's possible to evaluate and compare different algorithms' performances to establish the most robust and flexible one.

Chapter 5

Main algorithms concept

Both cameras have been tested in static conditions, hence with conveyor belt stopped, and in dynamical conditions, with objects moving on the conveyor belt. Thanks to post acquisition filters, static objects images can be enhanced in quality reducing presence of bad estimates and increasing reliability of depth values estimates. Images of moving objects, instead, need these filters to be deactivated to avoid alias problematics and wrong depth estimates. Without filters, these images depth estimates result more uncertain since they are more subjected to local depth oscillations.

Before introducing more complex and elaborated algorithms, some very simple but working ways of processing images are introduced.

5.1 Connecting contours method

To apply this method is necessary to have the camera set and fixed at a certain height. Height is a determinant factor to make the Connecting contours method able to segment easily objects. Setting the camera too far from the conveyor makes objects, which dimensions are small, very thin and little respect to the full image pixel dimensions. Their contours would result small and not well defined since low characterized in terms of pixels. Having bad and small contours lead to bad estimates of objects coordinates and orientation, for this reason it must be avoided. The idea of setting the camera far from the conveyor may arise observing local saturation images in Chapter 4 (Figure 4.13). Setting the camera far from its subjects, in fact, decreases the near infrared light source saturation effects on the camera field of view. This happens because the enlargement of camera FOV makes light source more attenuated and better distributed in the ambient. This camera setting is a way that can be followed to get not oversaturated images but carries on disadvantages linked to difficulties to detect defined and clear contours. In conclusion, the best camera setting for this method is: fixed,

near to the conveyor belt, as parallel to the conveyor belt as possible and possibly with conveyor belt color enhancing contrast with objects.

By suitable choosing a background able to enhance objects, like for example a dark conveyor belt with white objects, can be easier to extract contours from intensity/RGB images. This could be realized using a simple tuned Canny edge detector to find products contours. Canny edge detection respect to simple gradient based edge extraction methods allows to obtain more reliable and continuous contours respect to other techniques since it's based both on gradients and Non Maximum Suppression (NMS) along the gradients direction. It combines very powerful results in combination with low computational efficiency [12].

In general it's not granted to have completely closed contours generated and so it's necessary to close them before proceeding with eventual object identification and orientation. Connecting contour procedure can give some problems. First, it can attach objects that are not touching but are simply near as depicted in Figure 5.2. This is the major weakness of this method since it's able to manage only objects separated from the others. In case of near objects a contour fusion could happen generating one big blob from two separated but near objects. Second, contours can be closed wrongly giving objects shapes significantly different from true ones. Third, eventual spurious contours coming from the background may be merged with object contours during connecting operations.

These two last problems have to be avoided since wrong reconstructed contours may lead to wrong barycenter and orientation estimates.

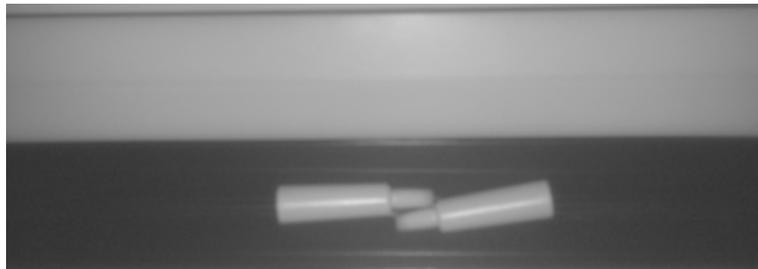


Figure 5.1: Intensity Image

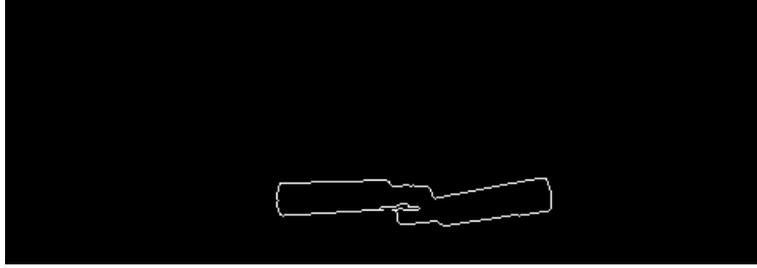


Figure 5.2: Connected contours Image

Based on the just detected contours area, it's necessary to filter out objects touching or overlapping in order to preserve only isolated and free to be picked objects. Objects interacting with others cannot be managed with simple blob analysis and, in an overimposition situation, it could not be easily established which one is on the top and which is on the bottom. Elementary OpenCv functions don't allow to separate easily objects in blobs so advanced shape matching methods have to be used. For this reason only single objects can be treated with this method and can be found by simply thresholding areas of all blobs found. By experimentally estimating one single object blob area is possible to understand if or not a blob includes one or more objects by comparing their values.

Now, calculating blobs moments, it's possible to extract from this filtered out objects their orientations and barycenter, to provide information at the robot arm for the pick up process. Barycenter coordinates (\bar{x}, \bar{y}) can be calculated through the following equations:

$$\bar{x} = \frac{m_{10}}{m_{00}} ; \bar{y} = \frac{m_{01}}{m_{00}} ; \quad (5.1)$$

where:

$$m_{ji} = \sum_{x,y} (\text{Image}(x,y) \cdot x^j \cdot y^i) \quad (5.2)$$

Thanks to Equation 5.3 and moments just calculated also the major axis orientation can be easily found.

$$\theta = -\frac{1}{2} * \text{atan} \left(\frac{2 * mu_{11}}{mu_{20} - mu_{02}} \right) \quad (5.3)$$

For the direction procedure third order central moment mu_{30} or third order normalized central moment nu_{30} are necessary (Equations 5.4 and 5.5). They have to be calculated after an horizontal reorientation of the object blob. Since all the objects are symmetric with respect to their major axis it's easier to establish which is their orientation direction when they are horizontally oriented. In case it's positive the direction result gives right otherwise gives left (Figure 5.3, 5.4).

$$\mathbf{mu}_{ji} = \sum_{x,y} (\text{Image}(x,y) \cdot (x - \bar{x})^j \cdot (y - \bar{y})^i) \quad (5.4)$$

$$\mathbf{nu}_{ji} = \frac{\mathbf{mu}_{ji}}{\mathbf{m}_{00}^{(i+j)/2+1}} \quad (5.5)$$

Figure 5.3: Object 1; $mu_{30} = 1.52953e+09$ Figure 5.4: Object 2; $mu_{30} = -1.36207e+09$

Method just described reduces the production efficiency in terms of ratio between grabbed and grabbable objects since all objects lying near or on other ones will be discarded a priori as false negatives. In addition, depth information are not used at all, that makes quite useless the usage of a 3D camera with respect to a 2D one.

5.1.1 Basler detail

In the case of Basler camera, objects' blobs could be directly provided by the application of a threshold at the confidence image. As already told in Chapter 4, Basler camera fatigue to map in depth dark objects and in particular dark conveyor belts. By using, then, a black conveyor belt, the confidence image will contain dark values in correspondence of the conveyor belt while brighter ones in the objects' areas. Separating brighter objects areas from darker conveyor parts with a threshold on the confidence image,

it's possible to obtain for free objects segmentation without passing by the contour detection procedure described in section before. This avoids all the connecting contour problems linked to wrong connections or contour closing and provides directly precise binary images containing blobs (Figure 5.5).

Now, applying blob filtering procedure to obtain only single products blobs, their own moments can be calculated to get for each object barycenter, orientation and direction (Figure 5.6). Since no object contour connection problem arises, this technique results more reliable but also more efficient. In fact, all these near objects that using the contour method were connected in one single blob, now are classified as different and separated blobs that are no more discarded.



Figure 5.5: Binary Thresholded Confidence Image

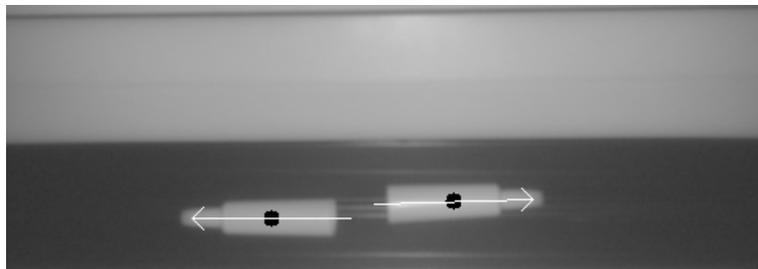


Figure 5.6: Intensity Image with barycenters (black points) and orientations (white arrows)

5.2 Background subtraction method

First of all must be told that this method is primarily thought for Basler images since they are more defined in terms of object depth shapes and positioning. In the major part of objects observed, this camera allows to map even objects' curvatures and details, except for possible wrong mappings as described in Chapter 4. RealSense camera, instead, doesn't provide such detailed depth images even if it's more globally accurate in depth estimates. Depth images, in fact, result less defined in terms of details and more color blurred in correspondence of most depth uncertain areas. For these reasons

from now on analysis and considerations done are referred only at Basler camera.

This method relies only on the assumption of having the camera fixed and approximately parallel respect to the conveyor belt. Conveyor belt distance from the camera can be changed provided that the objects continue to lie in the depth mapped range. There are a lot of examples in literature using this technique both with 2D images [13, 14] and with depth images [15, 16] to extract the features of the background in order to enhance and segment foreground components. In our case depth images are exploited to check if this method can work in our application field.

The specific method consists in collecting a background image of the conveyor belt without objects laying on it (Figure 5.7). Then, after collecting depth images of objects moving on the conveyor belt (Figure 5.8), the difference between this image and the background one provides another color image made of only positive difference values (Figure 5.9). Negative values obtained in the color difference are set automatically to zero values. In this way the multiple reflection effect described in Chapter 3 is eliminated since depth values deeper respect to ground level generates negative values in the difference image and are then mapped as black pixels.

Since color difference is based on the RGB channel difference then the RGB image obtained has to be converted into a grayscale image made of the sum of the three channels (Figure 5.10). In fact, the difference color image doesn't represent a depth variation scale because each channel carries information independently from the others about depth changes. A new image is created by summing the three color channels of the difference image. The color image, in fact, carries information in its channels, the higher the values in each channel the more the depth foreground image varies with respect to background. In this new image the difference between the two color images is mapped respectively from white to black into high to low depth differences. Now, on this grayscale image, we are interested in values showing changes of a certain magnitude. A thresholding filter is so applied to filter out low grey values coming from image disturbances and gray values showing that an object is effectively present (Figure 5.11).

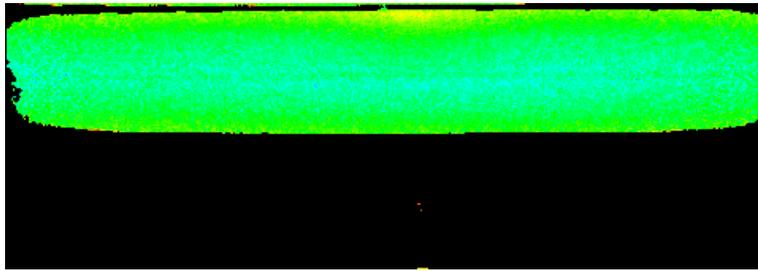


Figure 5.7: Background Sample Image

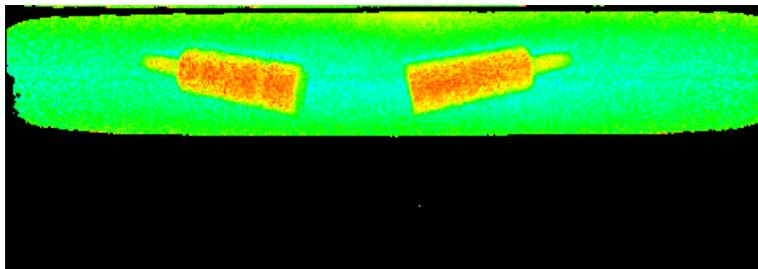


Figure 5.8: Foreground Depth Image

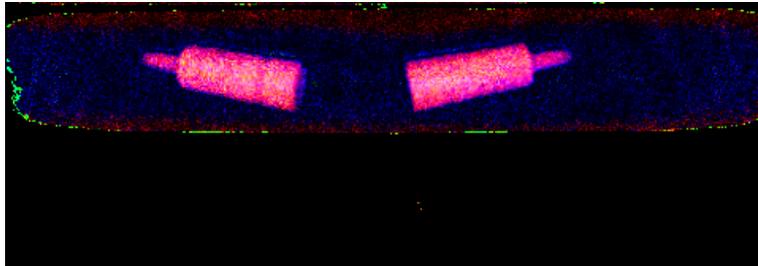


Figure 5.9: Foreground - Background Difference Image

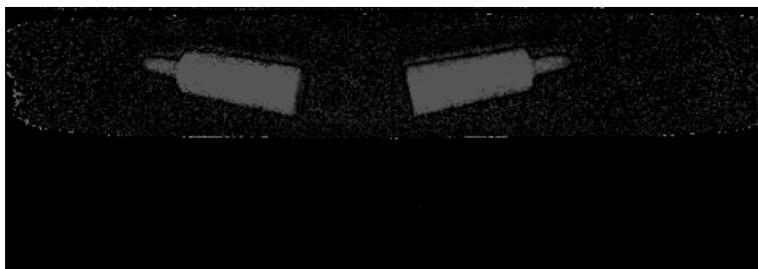


Figure 5.10: Grayscale Foreground - Background Difference Image



Figure 5.11: Binary Thresholded Gray scale Image

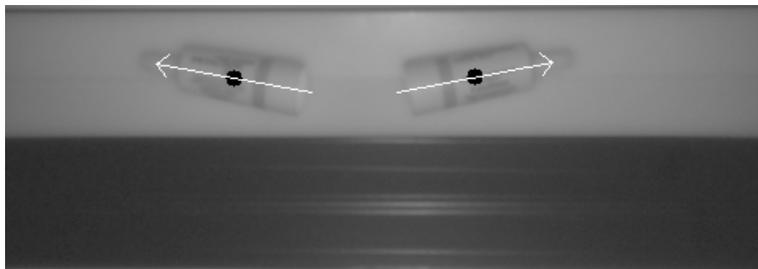


Figure 5.12: Intensity Image with barycenters (black points) and orientations (white arrows)

An advantage of this method is that it isn't required to have the camera perfectly aligned with respect to the conveyor belt since from background subtraction the conveyor height offset is cancelled and only the relative height changes are the ones enhanced. Another advantage of this method is that it's almost completely independent from intensity image. If, as in this example, the object color is the same of the conveyor, object detection techniques fatigue a lot to detect them since low contrast enhances their presence in the image.

Only single objects can be found since it's not possible by binary thresholded image to figure out objects' depth ordering when touching between themselves. It's necessary then to filter out, after the blob detection, blobs with an area too much different from the one of a single object. From now on single blobs features as barycentre, orientation and direction can be found simply using momentum techniques (Figure 5.12) in the same way described for the Connecting Contour technique.

Anyway camera precision is too low to grant, in absence of time and spatial filtering, a clear distinction between background and foreground for all kinds of objects examined. Since objects can be even thick less than a centimeter, it's difficult to mitigate disturbances' effect and perfectly segment objects. When the object's dimension is of the same magnitude of the depth estimation oscillations, it's difficult to select a right and effective threshold for the grayscale height difference image. If the thickness of all

considered objects would be high enough then this method could be successfully applied. For these reasons this method has to be abandoned and a more reliable one has to be searched.

For Intel Realsense a procedure of this type results not useful since, even filtering out objects blobs by background subtraction technique, the hard estimation uncertainties in objects contours doesn't provide a precise and detailed object shape.

Another fact to be considered is the non perfect coincidence between the object depth map and it's relative position in the RGB image. Even if alignment between RGB and depth image is done, hard uncertainties of depth estimates of the object generates wrong and imprecise object depth mapping. In other words, the depth map of the object created results not coincident with where the object truly is. In Figure 5.13, 5.14 an example of this phenomenon is shown. Depth and RGB images are aligned, by detecting the contour of an object (green color) and looking at the depth correspondent map, it's possible to observe that not all the depth blue map of the object is effectively inside of it. This means that this camera estimated badly depth around objects contours and this is justified by the fact High Density preset is used. Hence, basing analysis only on depth segmentation or on background subtraction in RealSense camera would drop the precision of required features extraction like barycentre and orientation.

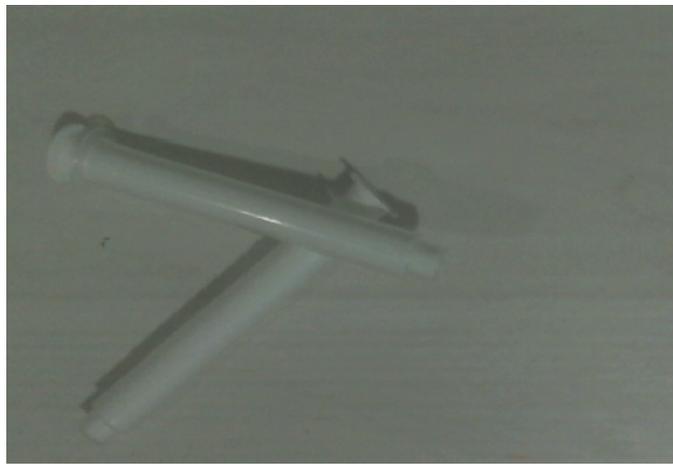


Figure 5.13: RGB Image

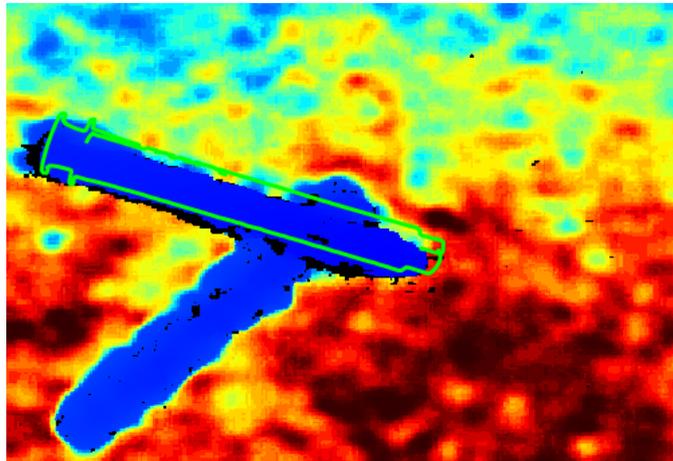


Figure 5.14: Depth Aligned Image with green detected object contour

Some advanced algorithm developed are able to manage more complex objects interaction detecting also free objects in overlapping relations through the additional usage of depth images. Their workflows will be now described in detail:

5.3 Search of the Object Model

In these analyses depth image and RGB/Mono images information are combined to exploit and demonstrate full advantages of 3D cameras. Instead, point clouds won't be considered. For simplicity and more clearness in the following chapter intensity and RGB images will be called field images.

5.3.1 Image prefiltering

First part of all three algorithms is equal and consists in filtering out the image parts where it's sure not to find some object. In this way the search procedure is executed only in parts in which it's possible to find objects, the search is sped up and the probability of finding objects where they are precisely placed is increased.

Two masks can be created as filters: first one consists in considering only the image part in which the conveyor belt passes so where the object could be eventually found. A simple rectangular cropping of the original depth and field images is applied restricting their Regions of Interest at the parts containing conveyor belt area.

The second mask is based on depth image information: only images parts in which depth is higher with respect to the conveyor belt depth value are conserved in the final image. In fact it's not said to have a perfectly horizontal conveyor belt and so some parts of it could be included in the search area even if they don't contain objects. In Figure 5.17 black areas highlighting the presence of objects are extracted to filter the RGB image. This image represents the Blue depth channel that doesn't carry information about objects depth since their depth map relies only on green and red colors.

Finally these regions are a bit dilated to help the shape based matching algorithm to find objects. Dilation is useful since in the final filtered field image some objects may result over cut, not showing their entire contour. In particular, in RealSense depth images, the depth blobs indicating the object are very approximative especially in their positioning and contours. In other words the blobs extracted don't match, in terms of coordinates, the RGB image object pixels. So it's fundamental to do a mask dilation before filtering out RGB images through masks. An example of filtering procedure is shown in Figures 5.15 - 5.18.

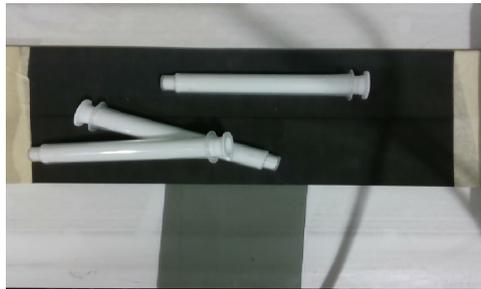


Figure 5.15: RGB Image

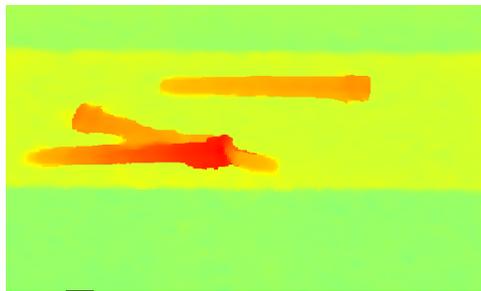


Figure 5.16: Depth Image

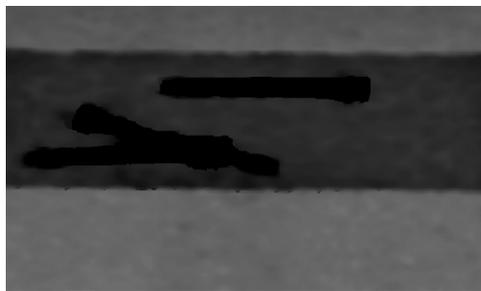


Figure 5.17: Blue channel Image



Figure 5.18: Filtered RGB Image

The procedures described for filtering the field images are not essential for the search algorithm to work. But they are essential to reduce timings of the search procedure and to increase precision against possible shape matching false positives. As explained in Chapter 3, an important requirement for algorithms is low execution time and the matching part of the algorithm takes the major part of the time needed for computation. It will be shown in Chapter 6 that reducing the image search area in object searching procedure decreases evidently the search time.

5.3.2 Matching

Search algorithms used come from a commercial library, called Halcon, because the equivalent function in Opencv is not very performant and so it's necessary to consider valid alternatives. From now on an explanation of main searching procedure stages will be provided.

There are two main different methods of searching objects in an image provided by Halcon libraries: shape-based matching and descriptor-based matching [17].

The first makes use of object contours to search for correspondances in images, the second tries to extract features from object textures and search for them in images granting also their localization and orientation. Since algorithm designed have to deal with textured and not textured objects and their orientation or rotation could vary, the second method results weaker and very object dependent respect to the first one. Another fact to be considered is that Basler doesn't see all objects' textures, as deeply explained in Chapter 4. So in the application considered the texture-based matching would be quite useless.

In conclusion, to provide the most general algorithm and to be flexible at all application cases, the texture-based matching has to be discarded in favor of the shape-based one.

The object matching configuration and workflow is divided in two parts:

The first one consists in creating a shape model through a sample image and a set of tuning parameters. Model creation requires the definition of an oriented ROI in the field image in which the object is contained. Thanks to some contrast and object dimension parameters the shape of the object is enhanced. In particular what has to be extracted are the object contours. By selecting in an accurate manner the rectangular ROI, the object orientation and barycenter come for free. In this way, along the search, as soon as the object is matched, precomputed shape barycenter and orientation are provided. It's important to choose the right and adapt image for the shape generation in which it's easy to extract object contours and there aren't

other distractor objects. In general an isolated object, near or in the field image center, is perfect. In this area light is homogeneous and perspective projection distortion of objects doesn't affect the model extracted. Perspective projection objects deformation is an important problem that has to be considered along the research and, as perspective projection rules specify, causes object deformation. In the next paragraph a method to deal with this problem, when searching for the shape model, is provided.

In the second part of the shape based algorithm, once a good contour model has been created, this contour can be searched in sample images. Several parameters can be specified, among them most important are: orientation search range, maximum number of matchings, matching score threshold, eventual search for scaled objects (smaller or larger), number of levels extracted from the model, search greediness, minimum contrast of search areas considered, shape deformation magnitude and the metric of shape-based search.

There are four types of metric: "use polarity", "ignore local polarity", "ignore global polarity" and "ignore color polarity". The fastest one is "use polarity" in which the polarity is observed, i.e., the points in the search image must show the same direction of the contrast as the corresponding points in the model. If, for example, the model is a bright object on a dark background, the object is found in the search images only if it is brighter than the background. Using instead "ignore global polarity" an object is recognized also if the direction of its contrast reverses. With "ignore local polarity" mode the object is found even if the contrast changes locally. Finally, if ignore color polarity is chosen, the matching can be performed in color images.

"Deformation factor" and "scaled search" are fundamental in order to find perspective distorted objects and 3D oriented objects. Deformation could happen in objects far from the image center and/or inclined. Deformation parameter magnitude specifies how different from the model the objects found could be. Since perspective deformation slightly changes objects shapes then this parameter grants flexibility along the search to find also these ones.

Another parameter that helps compensate distortion is the "scale factor". This parameter simply specifies if the model has to be scaled to bigger or smaller dimensions during the search. Some specific parameters set the angle and scale step of the model used during the search. Big values of these settings contribute to generating imprecise but faster search, vice versa too small values increases a lot the search process generating unnecessary too deep searches and unwanted increasing of the computation times.

With the help of the "optimization parameter" it's possible to reduce the number of the shape generated model points to speed up search. By increasing this parameter value, the model definition reduces and matching

quality could suffer from that. In particular as model point reduction is high as the matching could be bad. Point reduction may eliminate shape details causing, for example, approximative matchings or wrong oriented shape matching especially when the shape orientation is difficult to be determined since the model is very symmetric with respect to its major and minor axis.

”Search greediness” allows to decide how deep the search for object have to be done inside the image, the higher this value the more a complete search will be done and vice versa. A deep search increases the computation timings needed to conclude it, so a trade off has to be found between computation time and search quality [17, 18].

Once best trade off between search quality and computation timings have been found, search can be performed and results can be analyzed through three methods. A good search configuration should allow to find both isolated and occluded objects. The more objects are found the best the search algorithm is designed. As output the search provides a vector of scores that, of course, are major of the threshold set in search parameters. To each of these scores one position and one orientation angle are associated.

Position and angle orientation are sufficient to locate objects, thanks to a model shape rotation, in the exact position in which objects have been found.

Dynamic Case

Images kept with a standstill conveyor belt have important advantages with respect to images kept with moving one. First, moving objects need lower exposure times since motion blur has to be avoided. Images with motion blur could mislead the model search since generates not clear and sharp objects contour. Moreover some images can contain objects crossing the image borders going inside or passing away from the image field of view. The fact of having an object partially included in an image has to be managed carefully (Figure 5.19). In fact, these cases are very dangerous for the shape-based search algorithm since may lead to wrong matchings or simply wrong orientations of found objects. Since these error cases are not accepted at all in our applications, objects that are not fully inside collected images have to be excluded from the search regions.



Figure 5.19: Objects crossing Image borders

In particular to avoid searching objects crossing the image borders it's enough to exclude their regions from the search image. Halcon shape matching algorithm has the possibility to detect also objects crossing the image border but it's a weak method and sometimes could fail. It increases significantly also the computation times and this is a drawback in our framework.

The way chosen in which to proceed is the one of eliminating objects' areas if crossing the image border. In particular the procedure consists in creating a rectangular frame internal to the image contour and using it to exclude from search areas the objects having points in common with it. By intercepting each blob of the search region with this rectangular frame is known if the considered blob is inside the image or crossing its borders. If the area of the interception is different from zero then the object is crossing the image borders and the blob is subtracted from the RGB search region for the shape matching algorithm. If instead the resultant area is zero then the blob is inside the image and the search for the model in this area can be done. In Chapter 6 more details about timings and performances of this technique will be given. One important fact that will be enhanced is the high speed of computations that this method carries.

5.4 Derivative Analysis

After the object search is completed, it's possible to cut out the exact depth image information of each object by rototranslating the model contours with parameters specified by the shape-based matching algorithm output.

The core of this method consists in the extraction of pixel intensities from the grayscale depth image along the major axis line of the object's contours. By doing that, a model of the object's depth is created. Setting correctly depth range mapped by cameras is possible to obtain all the objects depth information mapped in one channel of the depth image (Figure 5.21). Choosing axis pixels is a reliable estimation of the object depth map since it's far from contours. Both Basler and Intel RealSense fatigue to map contours in depth. In particular Basler is affected by contours sinking due to multiple reflection phenomena (for major details see Chapter 4). Intel RealSense instead has the problem of giving very approximative mapping

around objects contours. Hence, keeping pixels on the axis of the object leads to avoiding this problem.

Example images are collected with Intel RealSense camera.

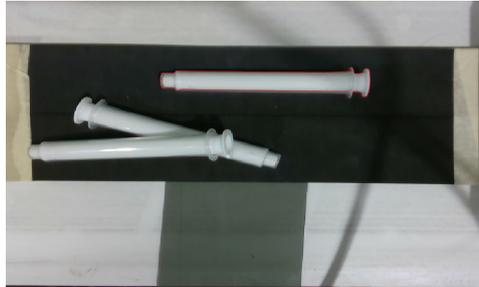


Figure 5.20: RGB Image with red detected object contour

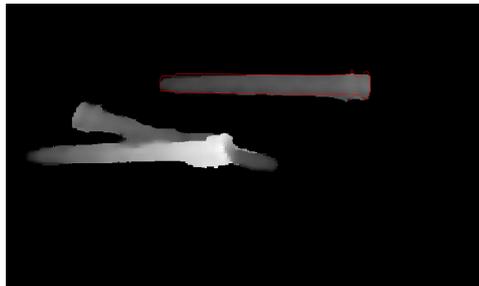


Figure 5.21: Red depth channel with red detected object contour



Figure 5.22: Free object Depth map

Once pixels intensity values inside the object axis have been collected, they can be mapped in a 2D depth function. The ground truth function expected to have, in case of a horizontal object, is a nearly horizontal line. How much the line is horizontal depends on how good the camera is placed parallel with respect to the conveyor belt and on the object shape. For example, having a plain object observed with a camera parallel to the conveyor belt would lead to an horizontal line. Instead if the camera is oriented respect to the plane this line will be inclined. Objects now considered are cylindrical syringes with the head little smaller with respect to the plunger part, so their depth map won't be perfectly horizontal.

Since pixel values are affected by intrinsic uncertainties and oscillation coming from technologies used, the resultant depth function results highly discontinuous as depicted in Figure 5.23. In blue is represented the 2D map of the object depth along its major axis, in red the derivative of this function. As it's possible to see, the derivative of the depth function is full of spikes since high discontinuities are present, so it's not very reliable.



Figure 5.23: 2D Object Depth map (blue) and its derivative (red)

This trend is clearly far from ground truth values previously described and makes necessary to filter out high frequencies components causing local unwanted discontinuities and derivative spikes. This filtering is realized by applying two 2D gaussian filters on the pixel's intensity array. The result produced (Figure 5.24) for the 2D maps is clearly more similar to the expected one.

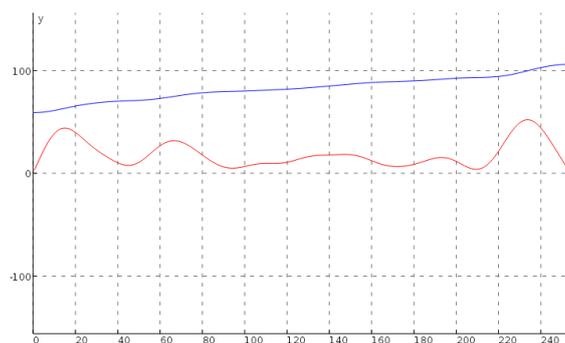


Figure 5.24: 2D Gaussian Smoothed Object Depth map (blue) and its derivative (red)

In order to check if gaussian filtering is too strong, the same procedure

has been applied to an object having another one on it. Figure 5.26 represents the 2D depth map obtained.

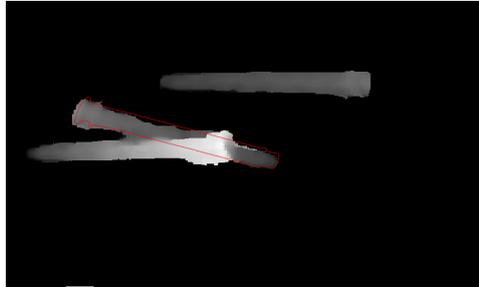


Figure 5.25: Red depth channel with red detected object contour

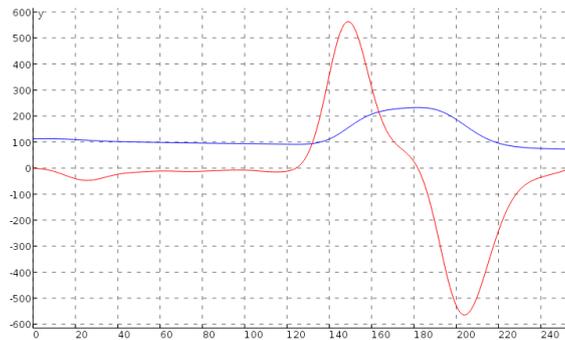


Figure 5.26: 2D Object Depth map (blue) and its derivative (red)

Filtering procedure doesn't compromise low frequencies components that are fundamental to detect eventual presence of objects on the examined one. It's evident that there's another object on the considered one, confirming that the gaussian filtering isn't too intense. This filters doesn't affect significantly the algorithm processing time taking competitively less than 0.1 ms of computation time but grants very effective and satisfying results. Having a good depth 2D map is necessary to create a better derivative of the depth function. This derivative function is the core part of the algorithm and, by studying this function, the presence of an object inside the considered one can be determined. Since the derivative function magnitude is small, a vertical dilation is done by multiplying the function values for a constant value. This dilation doesn't compromise the function characteristic but allows to study its behaviour better.

In presence of high depth discontinuities the derivative assumes very high or very low values. From the derivative plot (Figure 5.26) can be noticed that in correspondence of the object presence, highlighted by depth hump,

the derivative function reaches its maximum and minimum. By selecting the proper derivative threshold for the derivative magnitude the possibility to pick up an object can be decided. Derivative plot must be analyzed looking at its absolute value since both positive and negative peaks are indexes of overimposition presence.

Another consideration can be done about the strengths of the algorithm since even extreme overlapping cases could be handled right by this method. It's important to underline that there could be limit cases in which only one high positive or negative peak happens with an overlap. In Figure 5.29 is shown depth and derivative plots in the case in which the overlapping is located in the extreme part of an object. From these images it's evident that even a negative (but also positive) peak in the derivative plot means the object is not grabbable.

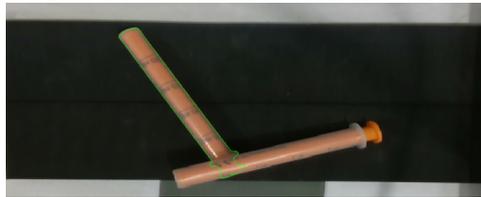


Figure 5.27: RGB Image with green detected object contour



Figure 5.28: Free object Depth map

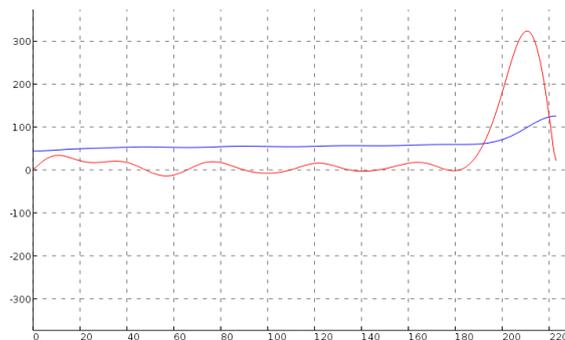


Figure 5.29: 2D Object Depth map (blue) and its derivative (red)

The magnitude threshold has to be established experimentally since it's strongly dependent on the objects thickness and geometry. In general with bigger objects the thresholds are easier to be established since depth variations are more large and evident and this reflects in high magnitude of derivative peaks values. Since false positives have to be avoided, the lowest derivative absolute value peak among all overlapping samples has to be kept as threshold. In this way more false negatives can happen in case of very ambiguous 2D depth derivative functions but no false positives at all will be found. Ambiguous plot happens especially in cases where objects are thin and local oscillation coming from disturbances are of a similar magnitude of a depth variation due to an overlap.

One very important thing that has to be said is that depth function plots considered until now represented only pixels belonging to the internal part of objects. The depth map representing a free object depth along its major axis should also contain high variations in the initial and final part (on the top and on the bottom of the object). In fact, as said several times, bad and uncertain contours estimates are given by both cameras. Furthermore, around objects is present the background so, even if the object depth map would be perfectly estimated, there should be a height step due to passage between object and background heights. These two problems reflect in hard depth variations on the extreme right and left of the plot. The real depth map is represented in Figure 5.30.



Figure 5.30: Complete 2D Object Depth map (blue) and its derivative (red)

These depth variations map into high magnitude peak derivatives. As described before, these high derivatives peaks would cause to discard all the objects.

The solution designed is able to cut out first and last depth map parts. The cutting dimension in each side is one fourth of the object width. This doesn't compromise the generality of the algorithm and makes the algorithm working in all possible overlap configurations. As demonstrated, with the

plot in Figure 5.29 even if the initial and final part of the objects is not considered, cases in which overlaps happen at objects' extremes are well managed. The depth discontinuity can be seen in the depth function and reflects in a single high derivative peak.

Another procedure has been adopted to make the depth map more reliable. When copying pixel depth intensities into the vector to be gaussian smoothed, instead of keeping directly the single current pixel value a mean of the upper and lower pixels is kept. Looking, for example, at Figure 5.22, for each pixel of the major axis the mean value of the current and the four upper and lower pixels (if different from zero) is assigned to the depth map vector. In this way it can be shown the depth map results less sensible to pixel local oscillations and more homogeneous, especially in Basler case.

Once an object has been classified as grabbable, it's necessary to tell also to the robot if the object is oriented in the z direction. This information is needed by the end effector vacuum cap to be always oriented perpendicular to the object. A solution for this task could be to consider standard deviation of the depths pixel vector just extracted. If the standard deviation is under a certain threshold then the object is horizontal, otherwise it's oriented. As already said this orientation task is required and done for objects already classified as free. Also mean value of depth intensities could be a valid alternative for the classification since oriented objects are higher respect to horizontal ones. A solution using mean values will be essential to solve this task in an example described in Chapter 6.

5.5 Common Area Method

Common Area Method relies on the features analysis of areas that two overlapping objects have in common. In particular this method works if and only if all the objects are found in the field image and consequently all common areas can be detected and analyzed.

Common area consists in the interceptions of each couple of objects present in the image. These areas can be obtained by scanning all possible pairs of objects found in the image and calculating their interception area. In other words the first method part consists in intercepting each object area with all the other objects areas one by one, analysing each area features. In case the interception is a null set, it means that the two objects considered don't touch and so don't overlap. Instead, in case the interception is not a null set (Figure 5.33), then further analysis could be done to establish which product lies on the other. This can be done by looking at the red common area in the grayscale depth map along the two objects orientation specified by the blue and yellow arrows as in Figure 5.33. Gray scale images provided come from Basler camera and, respect to Intel RealSense depth

maps, present darker values for near depth points and brighter values for far depth points.

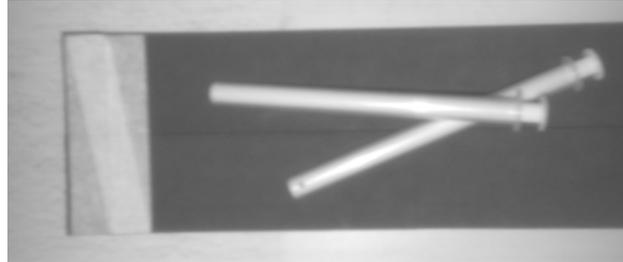


Figure 5.31: Intensity Image

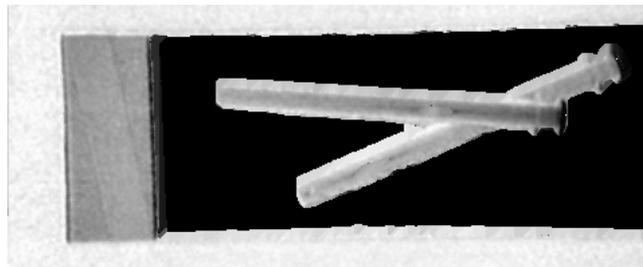


Figure 5.32: Depth Gray scale Image

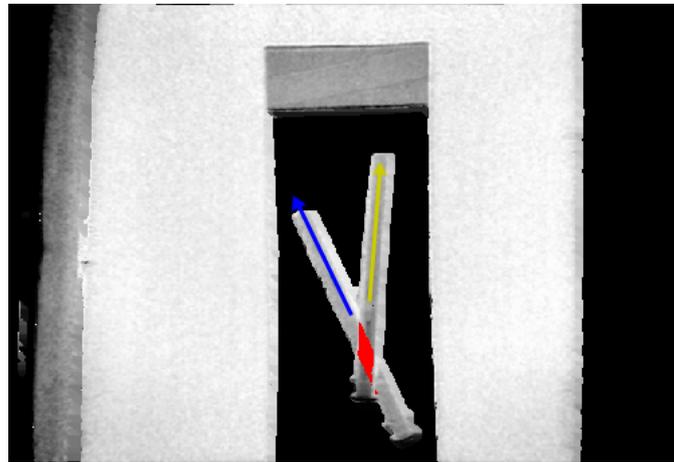


Figure 5.33: Depth Image with common area (red) and orientation arrows

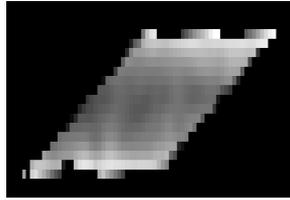


Figure 5.34: Aligned Common Area respect to yellow arrow

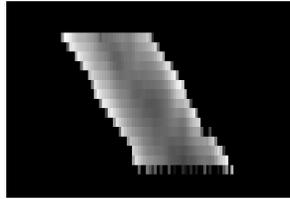


Figure 5.35: Aligned Common Area respect to blue arrow

By looking at this area reoriented in the two directions it's possible to establish the depth order of the two objects. The two reoriented grayscale common areas are depicted in Figures 5.34 and 5.35. Figure 5.34 shows an horizontal dark central line becoming brighter far from the image center, while Figure 5.35 shows horizontally the same behaviour: from bright to dark and then back to bright gray colors.

The first described pattern belongs to the higher object since the reoriented area has a continuous trend that reflects the depth trend along the yellow arrow direction.

The other area belongs to the lower one since it represents the overlapping area as seen by looking at the lower object along the blue arrow. An easy method to establish which object is up and which down is to consider the mean pixel horizontal value of the central pixel stripe in each of the two reoriented areas. The minor mean value will tell which object is up and the major which one is down. A plot of the two stripes values is represented in Figure 5.36 to show the clear shape difference of the two patterns. Since the blue plot assumes high almost stable values and the red one assumes increasing and decreasing values it's expected to have different mean values of these two functions. By experimental tests, it's demonstrated that considering mean values as indexes to establish depth hierarchy is strong and effective to counteract eventual local pixel estimation oscillations. Values represented in this plot are the 16 bit gray scale value coming from the inverted Common area images. The inversion, that maps dark colors into bright ones and vice versa, it's only done to have depth plot coherent with real objects depth curvatures.

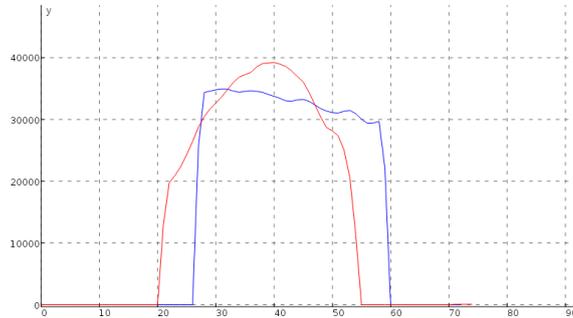


Figure 5.36: Depth map of central pixel stripes in Figures 5.31 and 5.32

UP/DOWN relationship is registered for each couple of objects having a common area. This leads to obtaining UP/DOWN information for each couple of objects. Once all this procedure is finished, the relations are examined to get the final products ordering.

The pick up ordering procedure consists in collecting UP/DOWN binary relationships in two vectors. Briefly the ordering methodology will be now explained.

The DOWN vector contains objects classified as under some other object. Analogously the UP vector contains objects classified as upper respect to another one. If an object is not contained at all in the DOWN vector it's, of course, a free object and so it's recorded as first in the pick up order. This object, in fact, as the definition of the DOWN vector tells, is not under any other object. After that, if some relationships involving this object are present, they're deleted from the UP and DOWN vector. This iterative procedure continues until all the objects order is established.

Considering the current problem requests, only free objects are interesting for the application. Then, the method can be stopped also as soon as all the top objects are found. In this way only the highest and free grabbable objects can be found. But this method, as demonstrated, could be more powerful and allows to establish global pick up ordering even if not strictly requested.

5.6 Multi-Layer Perceptrons (MLP) Method

An initial consideration about this method is that it works with at most two objects having an overlapping relationship. This means that the only overlapping cases treated by the following method are the ones in which one object is on another as is shown for example in Figure 5.37.

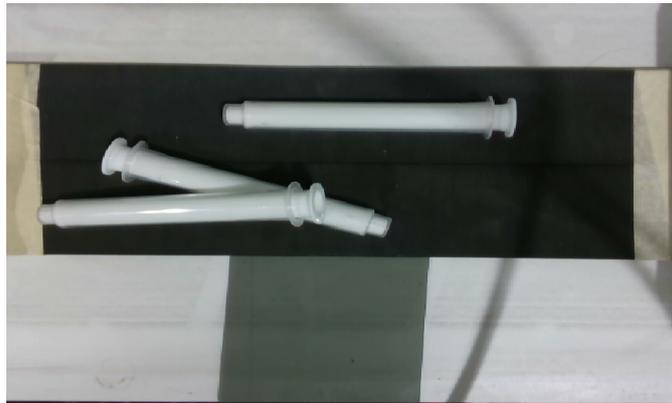


Figure 5.37: RGB Image

Some adaptations of this algorithm for more complex objects interactions will be described in the next Chapter even if interactions of more than two objects are from real machine cases.

In order to avoid having more objects near or on each other some particular types of conveyor belts can be used. Instead of being plain, they have like a V shape section, creating separated drawers on the conveyor belts. Thanks to V section and drawers division, objects are automatically separated in different drawers, limiting the number of objects near or on each other.

5.6.1 Working Principle

Generally a classifier is required when it's needed to assign an object to one of several available classes. The method used by technologies introduced is based on the separation of values assumed by different classes through lines, planes or in general hyperplanes. MLPs are a particular kind of neural network structured to be able to find these hyperplanes automatically by providing labeled train sets to them. Once these hyperplanes are determined through a training phase, new samples can be labelled. Based on the classes collocation and on the hyperplanes position just created, a classification label is assigned to each sample. In case classes are linearly separable then it's sufficient a neural network with only one layer. If instead relationships are more complex and classes are non linearly separable then more layers can be added to the net. These layers between the input and the output layers are called hidden layers. One hidden layer is sufficient to approximate any separating hypersurface and any output function with values in $[0,1]$ as long as the hidden layer has a sufficient number of processing units. In hidden units the activation function is the hyperbolic tangent function (Equation 5.5).

$$\tau_{anh}(x) = \frac{e^{2x} - 1}{e^{2x} + 1} \quad (5.6)$$

To find separating hypersurfaces for a classification, using a MLP neural net, the network weights have to be adjusted. This is called training phase and consists in inserting training data in the input layer and checking the correctness of the output layer that is responsible for the classification task. Based on the comparison between the training label assigned to data and the output of the network, the weights of the network are incrementally adjusted to minimize the error done in the estimation. Except for the input nodes, each node is a neuron that uses a nonlinear activation function. Halcon provides an advanced and very stable numeric algorithm that gives better results respect to the normally used back propagation algorithm. MLP can work for classification of general features, image segmentation, optical character recognition, least square fitting (regression) and classification problems with multiple independent logical attributes.

In most of the cases, it's recommended to use either the MLP, Support Vector Machine, Gaussian Mixture Models, or k-NN classifier, as these classification approaches are the most powerful and flexible ones. For completeness a description of all classification methods is done, then an explanation of main reasons for which MLP have been chosen will be provided:

- MLP classifiers have the characteristic of requiring more time to be trained, usually these nets are chosen in cases in which the training can be done in the offline phase and there are no constraints on timings. Anyway, in contrast with a long training time there is a fast classification time. MLP requires low memory to be run and a speed up of the network can be obtained by decreasing the number of hidden nodes and classes. For a general classification all types of features can be used to create the feature vector: region features, color or texture. In case it's necessary to add new additional training samples a new training of the net is required from scratch.
- The Support Vector Machine (SVM) classifier may give slightly higher classification quality with respect to other classifiers. It's slower in classification respect to MLP method but has a training time minor respect to it. In general is too slow to be used online and needs more memory than the MLP but less memory than the k-NN classifier. The memory required increases as the number of samples increases.
- The Gaussian Mixture Models (GMM) classifier grant both a fast training and classification, especially if the number of classes is low. It can be applied to classification cases that don't need high dimensional feature spaces. The feature space is the space created by the number of

features used to classify. It's very affine with novelty detection tasks. Novelty detection consists in detecting changes or defects of objects.

- Main advantages of k-NN classifier is to manage high dimensional features spaces and is very fast in the training phase. It's possible to add training data easily without retraining all the models. Anyway during classification this method is slower respect to MLP and requires more memory.
- The classifier based on a 2D histogram is suitable for the pixel-based image segmentation of two-channel images. It's a valid alternative for fast classification of two-channel images.
- The hyperbox and Euclidean classifiers can be used when feature vectors have low dimension, for example, when segmenting an image based on the pixel colors. They are very fast in particular in cases in which clusters are very compact. Compared to a LUT-accelerated classification using MLP, SVM, GMM, or k-NN, the memory requirements are lower and it's easier to represent the feature space. In fact, for image segmentation through look up tables is possible to speed up significantly computations. This procedure requires a slower offline phase and a higher quantity of memory respect to normal segmentation procedure. In addition the maximum number of channels of an image allowed to use this method is three.

Main classification methods characteristic can be summarized in the following table.

	MLP	SVM	GMM	k-NN
Training speed	slow	medium	fast	fast
Classification speed	fast	medium	fast	medium
Highest classification speed is reached for	low number of hidden nodes and classes	low number of support vectors	low number of classes	low number of training samples
Memory requirements (after re-moving the training samples from the classifier)	low	medium	low	high (trained samples cannot be removed)
Use of additional training data is possible without the need to retrain the whole classifier from scratch	no	not recommended	not recommended	yes
Suited for high dimensional feature spaces	yes	yes	no	yes
Suited for novelty detection	no	yes	yes	yes

Table 5.1: Comparison of the characteristics of the four classifiers MLP, SVM, GMM and k-NN. [19]

Among all described and examined methods, the best one to adopt is MLP method. It grants the necessary classification quality required by the current application together with fast online classification time. The training time, in fact, is not important since the training phase can be done offline and the training model can be saved to be only loaded when necessary. No novelty detection has to be performed or additional training datas have to be added in the designed algorithm. In conclusion, MLP networks are able to satisfy the classification requests of this problem with fast results and low memory consumption.

5.6.2 Method Description

The MLP network structure used is made of 3 layers. The input layer is made of three or one units depending on the input image number of channels.

If the image is color there will be three input nodes (neurons), if it's black and white only one input node will be present. The hidden layer is made of three units. And finally the output layer, that is responsible of the final classification decision of the network, is made of the 3 units that is the number of classes used.

Since mutually exclusive classification decisions are necessary in the output, the right activation function for the output layer is softmax (see Equation 5.7).

$$\mathbf{f}(x) = \frac{e^{x_j}}{\sum_{k=1}^n e^{x_k}} \quad (5.7)$$

with j the current node number and n number of input nodes

Before the training procedure, training datas are normalized and weights for the network nodes are randomly initialized. Normalization can be used if the mean and standard deviation of the feature vectors differs substantially from 0 and 1 (as in our case) or for data in which the components of the feature vectors are measured in different units.

Initially this method expects to load a MLP network to process depth images. The network keeps as input the depth image giving as output three different cluster regions depending on pixels colors. This model requires then to be flexible enough to classify well all images without being sensible to local colormap oscillations caused by intrinsic camera imprecisions in depth estimates.

In order to generate a classification model a small code routine is able to train a Multi-Layer Perceptrons network, save the trained network parameters and test it on other images.

First of all the procedure consists in choosing the suitable depth image to train the classifier network. To accomplish the train task it's necessary to manually select a rectangular area for each of the three classification labels. Those areas will be used later to train the network. In particular the image have to contain one clear example of overlapping to sample High label area, possibly one object free to pick to sample Low label area and finally a large part of background including the conveyor belt for the Background label. One example of image adapt to train the network is in Figure 5.38 where can also be seen the three rectangles selected to train the network containing pixels samples.

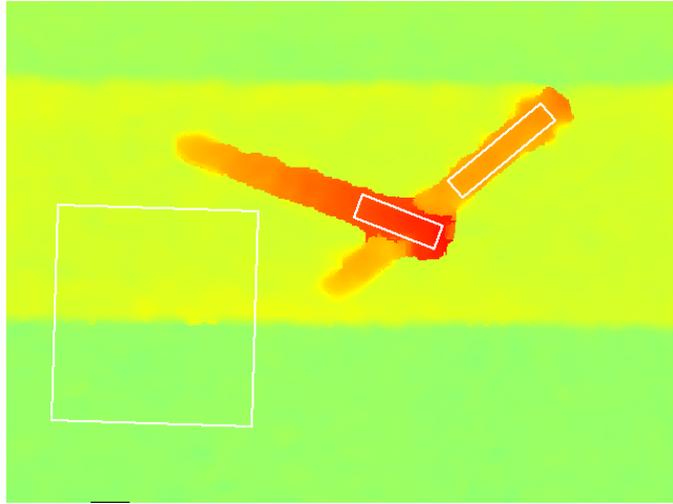


Figure 5.38: Training Samples Image

Finally a qualitative and not automatic testing procedure can be done by checking with other depth images classification results. If the training sets are well chosen and the network parameters are well dimensioned, the classification network should be able to classify different image areas pretty well.

The training procedure requires a number of maximum training iterations (set to 200) and two tolerance level that are responsible to stop the training before reaching the 200th iteration in case the training results and training weights becomes stable between two sequential training iterations ("WeightTolerance" and "ErrorTolerance").

Since depth image is subject to uncertainties and oscillations in values or, for example, the conveyor belt is not perfectly plain, the classification regions couldn't be perfectly right. Better results could be obtained using larger training sets or deeper and more complex networks. Anyway the classification obtained in the way described before is light, fast and strong enough to fulfill requirements of our application.

The MLP method, as described, consists of mapping color ranges into classes. This could also be done by filtering color ranges using in range filters or simple thresholding grayscale depth colormap. In other words, looking at color gradations in the depth images, color intervals can be separated by thresholding the image. This filtering procedure requires manual calibration that could be time consuming and sometimes complex to be done in each object case. In fact, it's necessary to finely establish limit range color between the labels. For this reason the MLP, that in this case acts as an automatic threshold generator, results preferable to be used instead of finding manually set points for thresholding the images.

In our context, in order to detect overlapping situation are necessary to

be used three classes: High, Low and Background. In Figure 5.38 are represented sample areas for these three classes. Background class is very useful to filter away image areas that don't contain information about objects. Then Low and High classes split the objects in two parts. Low classified parts represent the depth levels of free objects lying on the conveyor belt. The High classified parts instead are those parts that have a depth value lower (more near to the camera) respect to Low class depth range. Upper object parts in an overlapping relation will be classified as High. In Figure 5.40 the classification result is shown: Background is yellow, Low parts are violet and High parts are goldenrod.

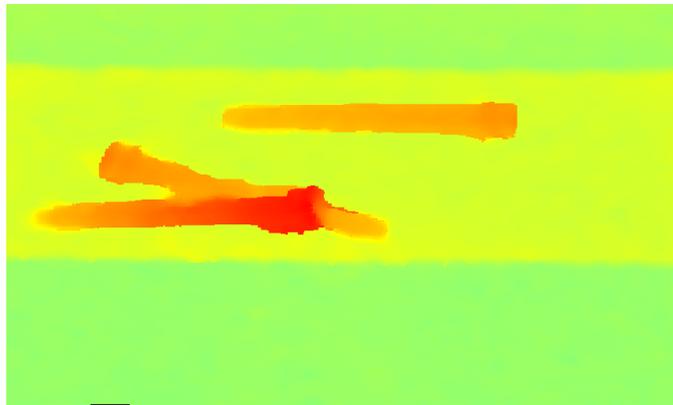


Figure 5.39: Depth Image correspondent to Figure 5.37



Figure 5.40: Classified Depth Image; Background (yellow), Low (violet) and High (goldenrod)

When an overlapping happens the High part is large and well defined but could happen to have local small and not expected areas labelled as High. If small zones are found, an initial filter is able to find and change their label to Low. Now the image is ready to be analyzed in search for objects

to be picked up or discarded. The decision, as is possible to imagine, is based on this new classification colormap obtained. High labels are colored in goldenrod, Low with violet and Background in yellow. The background region can be used as an alternative to the depth based filter discussed previously to exclude parts from the field image in which search the model.

Once the search image is defined by filtering out areas not containing objects, search can be started and the pose arrays of the found objects are obtained. After that, it's possible to locate their contours in the depth classification image created. Now for each object the area inside its contour in the classification image is inspected. In particular, what is determinant to decide if the object is free or not, is the amount of High (goldenrod) regions inside of its contour. Considering the assumption done at the beginning of this Paragraph about number of objects interacting in an overlap relation, the possible cases in which an object could be are three:

- Free object: this is the easiest case and consists in an object that has no objects on it, eventually could be near other free objects. It has to be picked up and its area should contain no goldenrod regions, only violet and at most a bit of yellow. (Figure 5.43)
- Object occluded: it's the case in which the object has another one on it and it has to be discarded. The contour area contains a small goldenrod area due to crossing of the higher component on it. The rest of the object is violet with a bit of yellow (Figure 5.41).
- Free inclined object: in this case the object is inclined since has another object under it, anyway it has to be picked up. The area of this object has to be almost all goldenrod, since the majority of it is over another one. The rest of it is violet with small parts of yellow (Figure 5.42).

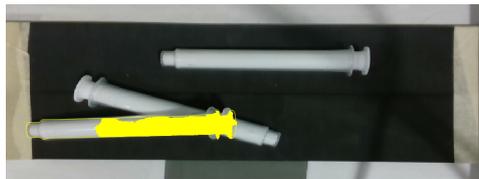


Figure 5.41: Free oriented object contour with High classified area inside of it (yellow)

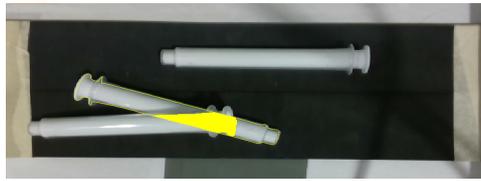


Figure 5.42: Occluded object contour with High classified area inside of it (yellow)

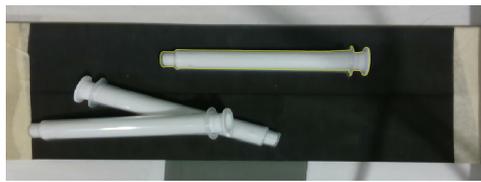


Figure 5.43: Free object contour without High classified area inside of it (yellow)

Hence if an object has no goldenrod parts it's for sure free. Since objects occluded and free oriented objects have evident different goldenrod area inside their contour, it's necessary to establish an area threshold for classifying the ones from the others. Thresholds clearly depend on how the SLM input image is structured and on the object's dimensions and shape.

As repeated several times false positives have to be avoided, so area thresholds have to be kept higher rather than the major value of goldenrod areas found in occluded objects.

Given the hypothesis of having at maximum two objects interacting, the decision about the z-orientation of an object comes for free when examining the goldenrod area inside the grabbable objects contours. If this area is large it means that the object is inclined (Figure 5.42), if it's small or absent it means that the object is horizontal and so not inclined (Figure 5.43).

5.7 Overall considerations

Common area method is a very powerful method since it's able to manage complex overlapping configuration with several objects giving possibility to establish also picking up order. Anyway this compromises the timings since it's necessary to consider all objects combination to establish relationship among them. This feature can be very useful in bin picking applications where all objects have to be picked up and timing constraints are not as strict as they are in the object moving application here considered. As will be said in the Results Chapter this method is difficult to be applied with Intel RealSense camera.

Instead, as already said, the Multilayer Perceptron Method can only manage two or at most three objects overlap and it doesn't provide the pick up order of them. MLP grants a robust classification technique that makes very easy and fast the classification of depth maps giving almost for free information about free or occluded objects.

Also the Derivative method is not designed for determining the order of objects to be picked up. Derivative method is based on the extraction of the object depth map along all the object axis. As will be shown in the next Chapter this is a weakness for this method since oversaturations and motion problems with Basler camera may give holes along the objects' axis.

In general each method behaves better or worse depending on the camera used and on the objects considered, the important result that will given in the next Chapter is that at least one of them is exploitable in every context.

Chapter 6

Results

This Chapter represents the core of the developed work and contains mainly all the test results, obtained and measured with different algorithms. Both Basler and Intel RealSense cameras are taken into account separately, specifying their main settings and acquisition frameworks. Settings are fundamental to get the best image quality for the algorithms used. For example, depth range mapping has to be finely tuned to get best performances for the research procedure. The research procedure, in fact, is based on the prefiltering of the RGB/intensity image using depth information. As said in Chapter 5, the filtering procedure for RealSense camera requires to have depth information of objects in one depth image channel. This is necessary in order to use one of the others to filter the RGB image for the search.

Algorithms presented in Chapter 5 will be tested with both cameras in specific static and dynamic working conditions. As will be seen, dynamic conditions, so having the conveyor belt moving, may worsen the performance of algorithms especially with Basler camera. Not all the three algorithms, as introduced before, behave well with both camera images and some of them will be unadoptable in certain applications.

No spatial or temporal filters should be used when dealing with moving objects. The resultant depth images, using these filters, could suffer from aliasing problems caused by the merging of information between current and past depth frames. The Spatial filter makes flatter and more homogeneous the depth estimates causing to lose depth details but granting to have less oscillating local depth values.

Another problem with which it's necessary to face up is the motion blur problem. Exposure times in both cameras have to be lowered to get better RGB/intensity image quality with sharp and clear contours. Otherwise the search algorithms may suffer from that since precise objects location or orientation may be uncertain due to objects contours are blurred.

Some considerations about conveyor belt types will be presented in order

to clarify the fact that its color, shape and light reflectivity characteristic influences deeply the results quality. Not all the conveyor belts behave in the same way especially with shape matching algorithms and with depth mapping technologies. Basler camera, for example, fatigues to map in depth dark objects and this plays a fundamental role in some algorithms presented. In addition, having a conveyor belt with colors in contrast with the objects makes easier the identification work for the shape matching algorithm.

Finally, some cases in which the algorithms presented are not applicable will be shown. In particular, analysis on overlapping conditions cannot be carried out since these objects are too thick and small. Anyway, some advantages granted by 3D cameras will be shown, starting from the reduction of the computation time along the research procedure up to the segmentation of the free isolated objects. This last advantage is mainly granted by the usage of black conveyor belts when considering the Basler camera. In conclusion, even in these cases, 3D cameras may make the difference with respect to 2D ones.

6.1 Conveyor Belt types

Before presenting results and comments about algorithms tests, some considerations are done for available conveyor belts types enhancing main problematic and advantages experimented with products. It's fundamental to consider this aspect of the working environment since both algorithms and camera performances can extremely change depending, basically, on their shape (flat or containing small drawers), their color (darker or brighter) and their light behaviour (lucid or opaque). Finally one conveyor belt type will be chosen for more exhaustive tests on algorithms performances in static and dynamic conditions.

6.1.1 White lucid plain conveyor belt

Choosing a white lucid conveyor belt makes difficult to detect objects with both cameras. Since the major parts of objects are white, the research procedure is affected badly for the fact that no contrast provided by the background color leads to not having clear and sharp edges. For the same reason, it's very hard also to generate object models since contours are not evident and detectable. Having objects difficult to be found reflects in having high greediness values in the search procedure. In this way a deeper research is done and less false positives arise. Using high greediness values also counteract eventual wrong orientation problems but this approach has the disadvantage of increasing remarkably the search computation times.

An exception is when, for example, colored syringes are considered. In this case the contrast with white background is higher. This allows to have

lower greediness values, so reduced search timings. The generation of the shape model and its research result easier and less wrong orientation problems arise.

With Intel Realsense camera, the presence of shadows may compromise results of the shape matching algorithm. Since RGB camera is sensitive to ambient light, shadows of objects are present in the image. The search algorithm searches for objects contours based on contrast variation in the RGB image. Since shadows have more contrast respect to objects on white conveyors, some matches of the model along shadows may happen. The presence of shape matching false positives have to be avoided since it's not ammissed by the application to provide wrong object positionings. In literature are present some shadows removing techniques but they would enlarge the computational effort and increase timing of the designed algorithms. Looking at Figure 6.1, the shadow problem described before with a wrong shape matching result is shown. The positioning of the object detected is highlighted by green contour. The shadow near the object tricked the search algorithm leading to a completely wrong matching.

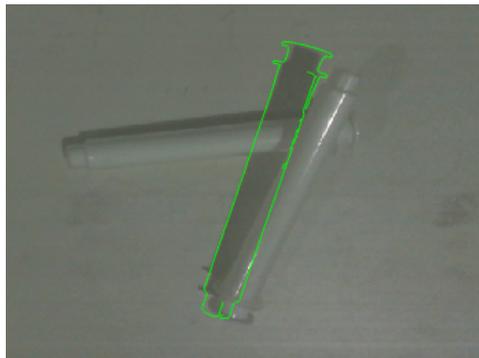


Figure 6.1: Wrong shape matching result due to shadow presence

Basler camera, on the contrary respect to RealSense one, provides diffuse illumination of the environment and, for this reason, doesn't present evident problems linked to shadows generation.

Anyway, white lucid conveyor presents additional drawbacks in case of Basler camera since it produces high reflective areas in the image center, causing circular saturations in the intensity image and consequent black not mapped areas in depth image. As already told in Chapter 4, high reflective lucid materials have to be avoided with this camera since they may act as a mirror with incident light generating saturation areas. On the other hand the central image part is the more reliable in terms of estimates and, for sure, objects passing under the camera have to pass from the center of its field of view. In conclusion, this conveyor belt can't be used with this kind of

camera since it's not admissible to have huge saturation areas in the central part of the image.

Another problem typical of plain conveyors is the rolling of objects. Since the conveyor is moved by an alternated motion, objects on it may be affected by acceleration and deceleration phases. In particular products with oval or circular section may start rolling and moving on the conveyor when variation of velocity happens. This phenomenon causes the product to roll and move on the conveyor belt. The positioning of objects estimated on the conveyor belt may then vary in the period between the moment in which the image of the object is kept and the moment in which objects lays under the robot arm. These changes of positions in time are not accepted since the robot won't be able to pick up objects. In order to solve this problem different types of conveyor belts have to be used with different sections like the V one that will be explained in next Paragraphs.

6.1.2 Dark opaque plain conveyor belt

This conveyor is made of opaque material so avoid high reflection problems with respect to the white lucid one but it maintains the problem of rolling away of objects.

Realsense behaves very well with that conveyor since shadows are almost invisible with a black background and problems of wrong detection described in the previous paragraph are eliminated (Figure 6.1). Without a dark background the identification problem was infeasible since the risks of having shape matching false positives or wrong shape orientations was too high. Now objects contrast is enhanced making objects research easier and faster with very low probability to detect shape matching false positives or false orientations. Also transparent parts are more evident with black background and a perfect shape can be easily extracted for the search part. Only dark objects are difficult to be detected and other methods have to be studied to easily locate them, maybe also using depth information.

Basler camera presents, instead, a nice behaviour with black background since it absorbs almost all the projected light and it's very difficult to be mapped in depth. High oscillations of depth estimates are localized in presence of the conveyor generating apparently an almost unusable depth map. Considering black areas in the confidence map they result, as expected, very low in value. As introduced in Paragraph 4.1.1, this phenomenon is due to the fact dark components absorb a lot of light and by well illuminating objects that are bright, dark ones are not sufficiently lightened. Since low quantities of light come back from dark parts, their depth estimation is bad. To check these considerations it's enough to increase exposure and see that the confidence image of the black conveyor becomes brighter in dark parts and as a consequence their depth estimation becomes more stable and reli-

able. By filtering out through a confidence map threshold darker parts, it's possible to obtain a depth map only containing objects depth information. Thanks to this result the search areas are generated for free reducing the search regions for the shape search algorithm.

Another problem that is still present, but manageable in a better way, is multiple reflection problems which causes bad depth mapping near edges. Thanks to black background, rays that are reflected by objects and should bounce on the ground before coming back to the camera sensor are absorbed by the background itself and so the sinking of contours in depth maps can be reduced evidently.

6.1.3 Lucid V section conveyor belt

This kind of conveyor belt is not plain but, as the name suggests, has a V section in each segment. It has one main advantage linked to object separation since, when objects lie on it, its shape automatically divides them. For example, objects like syringes or small pipettes moving on this conveyor are automatically separated thanks to the V shape drawers. This grants objects separation for the pick up process, avoid rolling away of objects during acceleration and deceleration of conveyor and reduce probability of having interaction of more than two objects in the same V drawer. Object separation is fundamental since during the pick up process of the robot arm, the object caught could be in contact with a near one. In this case, even if the other object is classified as free, the robot arm has to discard it since during the previous pick up process it could have been moved slightly from its position. This happens only when objects are both free but they are touching in some parts. In this way, free objects are discarded and the efficiency index is reduced. The V shape conveyor may avoid or reduce this problem since it intrinsically divides objects in drawers.

As introduced in Paragraph 6.1.1, the rolling out problem could be very critical in certain situations especially when the objects are conic or cylindrical and so easily start to roll. With an alternate movement of the conveyor belt, even this kind of conveyor belt may generate local objects oscillations. These local oscillations may modify objects positioning. In general an alternated motion has to be avoided in favor of continuous motion.

Having a lot of products interacting and eventually overlapping, could cause the shape matching algorithm not to detect all the objects. It may happen that hardly occluded objects are not found during the research since only partial object contours are visible. The orientation procedure becomes difficult if hard occlusion happens in the top or bottom parts of the objects.

An advantage of using this conveyor is that objects interactions are reduced, especially the overlapping of them. The V shape of the conveyor naturally divides objects inside the drawers. In this way it's easier to detect all the objects for the reason that they are isolated or collected in very

small groups. Since the Multi Layer Perceptron method requires few objects interacting when an overlap happens, this conveyor can be a very good choice with respect to a plain one for this particular type of algorithm. Some considerations about the effective application of the MLP method with this particular kind of conveyor will be done ahead.

One white and one dark conveyor of this type have been tested. White one still presents the drawbacks described in the plain white conveyor paragraph: with Basler camera, high reflections saturate the depth image creating holes and also the problem of multiple laser reflections arises, causing objects to collapse under the plane. In next paragraphs, a filling technique to fix local saturation problems reconstructing depth images will be provided. The bad mapping problem linked to multiple reflections may be overcome by using the Background Subtraction method described in Chapter 5.

About Intel Realsense camera, the shadows problem, described in Paragraph 6.1.1, of having shape based matching false positives make unacceptable results even with this white conveyor.

The dark one is made of reflective material too and presents small local reflections problems that may confuse shape-based matching algorithms when using the RealSense camera. In particular objects are found in positions in which they are located but sometimes partial wrong or inaccurate results may arise due to local reflections of light on the conveyor around the object location. Considering the Basler camera, instead, the conveyor is badly depth mapped since black absorbs light. The same phenomenon described with plain dark conveyors happens and can be considered as a big advantage since objects can be segmented for free. It's fundamental to remember that the effect of invisibility of dark objects happens with low exposure times. As exposure is increased as also dark objects are illuminated and start to be mapped in depth with higher confidence.

A usage of this kind of conveyor will result optimal if it would be made of opaque dark material. In that case reflection problems are avoided, Basler associates very low confidence values to dark background enhancing only areas in which objects have to be searched. This allows to reduce the search area and precompute isolated objects orientations.

Instead, IntelRealsense would provide images without unwanted reflections and shadows. The shape based search algorithm would work very well since no local distracting reflections would arise. V shaped conveyor also avoids rolling away of products and automatically divides and separates them in small groups.

6.2 Main Settings

For all the tests and results described from now on, the conveyor belt used is the black plain one since grants best results for both cameras and reduces at minimum problems of shape matching and light reflection giving best conditions for identification. The only problem, that will be remarked ahead, is that this kind of conveyor is plain and doesn't have, for example, a V section. The rolling away problem causing the possibility for objects to roll out from the conveyor during its movements is still present. In addition the V section grants to automatically divide objects inside drawers reducing the maximum number of objects interacting. This aspect can be particularly useful for the hypothesis required to adopt the Multi Layer Perceptron method that needs at most two objects interacting.

Conveyor belt variations

Except for the Multi-Layer Perceptron method, the algorithm used are insensitive to the section shape of the conveyor belt. It's not required, then, to change the algorithm structure if the conveyor belt changes in type.

Connecting contours method doesn't rely on depth image information and the Background subtraction method uses depth difference. As a consequence, Background subtraction is sensible only to the relative depth changes. Hence, the difference between depth image and foreground image makes the resultant image unaffected by the background depth map shape.

Derivative and Common Area methods are based on a simply shape matching on the RGB/mono images. Therefore, having parts of the conveyor belt at the same height as the products (like in the V shape case) or changing the conveyor shape without modifying the algorithm does not influence the algorithm workflow in any way. Maybe the image filtering part is affected by the conveyor belt shape since the search regions are created by filtering depth ranges. In case of using a V shape conveyor belt the search area may result larger. This phenomenon happens since the filtering is based on the depth levels and having parts of the conveyor at the same height of the object will cause to have them included in the search area. Anyway, except for this change in the filtering of the search area, the final results don't change.

Instead, talking about the MLP method, the usage of a not plain conveyor may compromise the results. In fact, MLP is based on a color labeling that associates a label to a pixel depending on the range in which its color falls. Using a conveyor not plain may cause some objects to be free but inclined. Their color map could be different from an object not inclined lying on a plain conveyor and, as a consequence, the labelling procedure could be partially wrong. This bad mapping phenomenon depends also on the

object considered, a thick object would have less problems with respect to a thin one. In fact, the labelling color ranges are larger and more robust in thick objects since the color variations in an overlapping situation are more evident. So, even if an object is a bit inclined but free, the classification map could not significantly change respect to the map of a not inclined free one.

Of course this method loses reliability in case of its usage with V shape conveyor belt, some modification to it has to be studied to make it adaptable.

Shape matching metric

The metric used to generate and search the model is "polarity" one since neither local nor global contrast variation in objects images is expected. This assumption makes more computationally lighter the search algorithm respect to the usage of other metrics like "ignore local polarity" or "ignore global polarity". "Polarity" mode, in fact, is the fastest modality and helps to reduce the computation times as requested by the task specifications. It's not flexible to contrast changes in search images but this feature is not required in current applications. In particular, tests done with several images of several products showed that the "polarity" mode is the best choice in terms of performances and computation times. In particular, with high contrast between objects and foreground, the model generation and the identification quality are very high and the research does not require deep searches using high values of greediness. It's not necessary, then, to make deeper and longer searches using parameters different from the "polarity" one.

Cameras Distance

The distance from the conveyor at which cameras are set is about 30 cm, this allow to have a large enough field of view to clearly see objects including cases of groups of them. This is the near extreme distance from which the Basler camera can be used. Higher distances could be used but it's not suggested since as distance increases as objects becomes smaller and smaller causing difficulties in their research and identification. It's important to remember that objects considered are in the dimension order of a pen, going far from them would make them too small to be well detected by shape matching algorithms or contour detection techniques. The closer is taken the image of an object, the more detailed its definition will be in terms of pixels.

Vice versa going closer than 30 cm to the conveyor belt with cameras would reduce too much their field of view. Then, having a group of objects, it could not be granted to have all objects included in the image and in these cases no analysis can be done. If all the objects are crossing borders of the image, the region containing these objects would be always subtracted

from the search image region and, as a consequence, no object detection will be done. Remember that this region subtraction method was introduced to avoid the increase of computation time due to object research and because object detection fails easily in detecting objects partially included in the image. More details about this procedure are given in Paragraph 5.3.2 .

Another reason for which Basler cannot be used nearer the conveyor belt is that its working range is at minimum 30-50 cm so lower distances could cause problems in focusing objects or mapping depth.

Offline computations

All variable allocations and definition times, including their initialization, are not considered since they can be done offline in the machine initialization phase. Also timings to train or load MLP models and load the shape model to be searched from the shape-based matching algorithm are neglected for the same reason. These procedures can be done once for all before the starting of the machine, after the model format to be processed has been chosen and the machine parts are initialized.

It's not told that the same machine could process different types of products, for example, syringes. In this case for each product will exist a certain camera preset, a certain object model, eventually a MLP model and dedicated algorithms parameters. In order to change all the parameters only the object format needs to be changed.

Computational timing

About algorithms and procedures timings specified in further analysis, some important considerations have to be done.

The execution timings measured are related to a commercial computer without any typology of real time execution. In particular the processor model used is Intel i7-8665U. These timings are then indicative since other processes executed in the operative system may require the preemption of the current algorithm delaying its execution. A true and reliable measurement of timings should be done on an industrial PLC to check effectively how much time is required to execute the algorithms and to know precise advantages in timings. Anyway, qualitative considerations on execution timings of some parts of code could be done in order to enhance advantages in adopting or not adopting certain solutions. For example, it will be shown that filtering the search image may decrease evidently the time execution of the algorithms. This is a very important result since it can be applied only when using depth information and, so, a 3D camera.

6.2.1 Intel RealSense D435

Acquisition parameters like resolution, framerate, exposure time, histogram equalization, depth clamping values, including also post processing filters and presets can be set before acquiring images from RealSense Viewer software. In this way parameters are set in a static manner and they can't be changed during acquisition procedure.

In alternative, they can be decided and changed along acquisition procedure integrating the acquisition C++ program with libRealSense functions and/or json files. These libraries are very well documented and complete. They allow to change all camera parameters without passing from the RealSense Viewer software. Json files consist simply in lists of parameters names and correspondent values that can be imported and applied to the camera once for all.

RealSense viewer provides the acquisition feature of single frames or frames sequences but doesn't allow to align them and save directly after acquisition. It's always necessary a post elaboration of these frames in order to align them through the alignment process described before. Alignment is based on several camera parameters, it cannot be achieved only having depth and RGB frames. It keeps into account images resolutions but also baseline and disparity values so on the Disparity shift used. This last parameter is very useful for near objects applications since it allows to reduce the minimum mapped distance. The scope of this setting is to reduce the minimum distance at which the camera can map depth data and relies on the reduction of the disparity distance among images. Increasing the Disparity Shift the maximum distance mapped in depth by the camera decreases evidently. Anyway, at 30 cm distance with respect to objects. the usage of this parameter is not required.

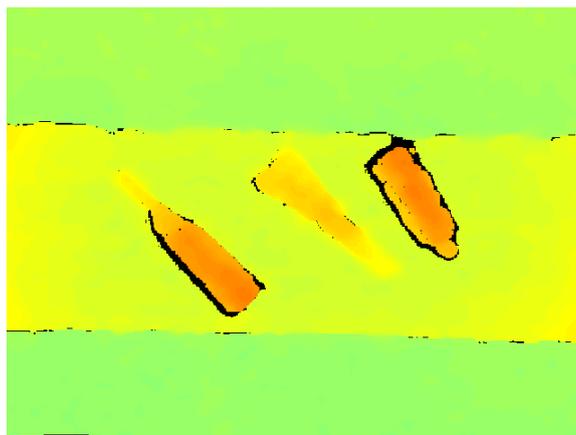


Figure 6.2: Depth map without filling procedure

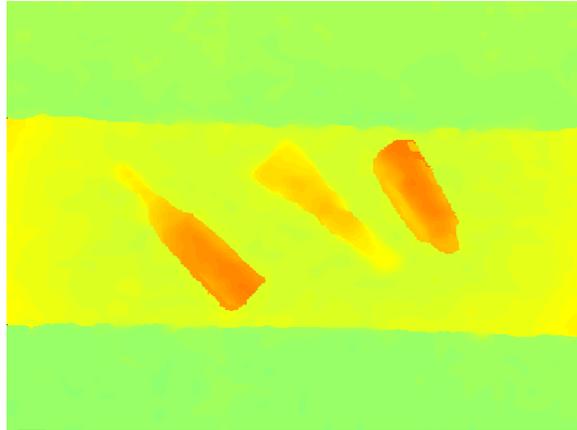


Figure 6.3: Depth map with filling procedure

Since High Density preset is used, depth map images have high density concentration of pixels like in Figure 6.2 . Anyway both MLP and Derivative methods may have problems in using depth maps with holes inside of them causing high discontinuities in the depth maps. Two ways to solve this problem have been used: decreasing the "DS Second Peak Threshold" and activating the Filling Holes post processing procedure. In this way the depth map image can be completely filled by depth estimates. Even if these estimates are very low in reliability, it will be shown that having a Filling Rate value at 100 % makes safer and easier to apply advanced algorithms procedures (Figure 6.3). The Filling Holes post processing filter is the only one adopted in these tests. Spatial and Temporal Filters are not used since depth estimates are stable and have no high oscillations. Spatial filters contribute to make homogeneous and flat the depth estimates compromising eventual discontinuities useful for the application of Derivative method. Temporal filter, even if could be useful to have more precise depth estimates for objects, creates aliases with moving objects that have to be avoided since they ruin overall depth estimates.

The depth mapping set up used is: without the equalization histogram mode, with jet mapping colormap and with min and max range set respectively between 0.2 m and 0.32 m. A too strict mapping range would lead to having images too much detailed in terms of depth precision. In particular with Histogram Equalization the depth image color results highly locally oscillating. The depth estimates result very difficult to be used since high local depth discontinuities are mapped in high local derivatives causing the depth plots to be difficult to be used especially with the Derivative method (spotted effect)(Figure 5.14). For this reason the depth range has been increased.

Too high depth ranges, on the other hand, cause too much quantization in depth estimates making the depth map too homogenous and without local

depth variations.

The depth map has also to grant that one channel gives direct information about the depth segmentation as required by the search algorithm. In the example presented in Paragraph 5.3.1 and in further analysis, the red and blue channels don't contain information about depth and the green one contains objects' depth map in grayscale. In this way from red and blue images a simple binary threshold and a circular dilation (with radius between 7 and 10) allows to have directly the search regions.

In conclusion the best way to approach parameter acquisition and settings problems is to manage all in a common C++ program where camera settings are decided through a json file or libRealSense functions. In this way also image sequences can be handled easily since, after acquisition, alignment comes for free, thanks to the fact all camera settings are carried on by the frame grabbed information. 640 x 480 aligned images sequences are collected and ready to be processed by object identification algorithms.

Of course the method just described is the one adopted in this work to collect both single images and/or sequences. Anyway it's important to say that the RealSense Viewer results very useful during the initial testing procedure. In order to know if an object is affine and well mapped with stereo camera procedures, the Viewer is fundamental since all camera parameters can be quickly and easily set. All preliminar results shown in Chapter 4 are done using mainly the Viewer application that remarkably speeds up the discard procedure for objects not well depth mapped by this camera. In order to have sequence acquisition and alignment for the final robot arm application a C++ program has been created.

Intel provides, with its RealSense viewer program, some demo codes showing some basic functions provided from the libRealsense camera integrated libraries. Among these functionalities a fundamental one for the application studied allows to pass from a depth pixel color data to a distance in metres. This comes very useful in robot picking applications since it's necessary to know at which height is the object to be grabbed. Once the positioning of an object to be picked up is determined, so its orientation and barycenter are known, in order to provide to the robot arm a space positioning of the object, it's necessary also the depth information of this object.

The second very important functionality deals with the depth-RGB alignment. As already said, thanks to this alignment procedure, it's possible to get depth and RGB images perfectly aligned in terms of coordinate systems. In this way referring to a certain pixel in depth image is possible to get directly its depth value in depth image at same coordinates.

Thanks to depth information and alignment is possible finally to carry out

the depth information of the object found and the (x, y, z, α) object coordinates can be determined.

In images shown the High Density preset has been used, carrying mapping errors especially in edges or non textured areas. As said before, this preset has low matching threshold value leading to uncertain estimates. In particular, from a detailed analysis of aligned images proposed (for example Figure 6.3), the orange-red areas enlightening objects in depth map doesn't correspond or map precisely objects contours. This is also additionally compromised by the object's depth images filling procedure. Furthermore, if the two images are aligned, these colored areas don't perfectly include the correspondent objects in the RGB image. This means that a pure segmentation starting from the depth image is not possible in the RealSense case since segmenting orange-red color won't lead to segment perfectly object and consequently to map in a good way their position in space for the robot arm.

6.2.2 Basler blaze 101

Basler provides for the Basler camera a very useful viewer program that allows to observe directly main parameters changes on the collected images. In this way a tuning of parameters can be easily done for having best output image quality. Main camera parameters to be set are exposure time, depth range, operation mode, spatial and temporal filter settings. The higher the exposure time the more stable are depth estimation images, on the other hand intensity images are more bright and with lucid objects and the probability of having large holes in depth map increases.

It's important to remember that, with Basler camera, once parameters are tuned for a certain application, the parameters don't need to be changed if light conditions vary. In fact the Basler camera is unsensible to ambient light conditions and is provided with band pass filters that makes the camera sensible only to its own infrared light projected.

Along with the viewer program some code demos are provided giving the possibility to extract Basler raw images data and process them in order to get intensity image, depth maps (color and gray scale) and point cloud files. There are also already coded methods to pass from the camera measured distance in each pixel to the relative color in the depth image. Depending on the range of scaling chosen the color associated to a certain distance varies consequently.

An interesting feature present in the Basler viewer, that was not so easy to be used in Intel RealSense camera, is the possibility to collect image sequences. The higher the quantity of files saved for each frame grabbed the less will be the frames saved in a certain period of time. For this reason,

only essential images have to be collected in order to have more frames saved in a certain time span. Mainly, the images needed and chosen to save in a sequence are: intensity image, confidence image and depth image (both grayscale and color). Even if the FPS given by the camera are higher, it's impossible to save all frames images since memory access and saving process require too much time.

The main difference between RealSense and Basler Viewer is that the Basler one saves a sequence of images directly in a format easy to be processed (jpg, TIFF, png...). RealSense, instead, compresses images collected in one file, making difficult to have a fast and easy extraction and processing of them.

Like in the case of RealSense camera, a program able to collect and save image sequences has been developed to provide an acquisition solution useful to catch images in a real machine. This has been done also to acquire confidence with Basler camera library and to test its flexibility in terms of automatic camera parameters settings. All parameters that can be changed by the Viewer may be set by using the libraries directives.

The libraries used and tested, along with images provided in this work are related to a not completely stable and definitive firmware. The Basler camera firmware was and is still under development so tests on the library stability was essential with this particular camera.

6.3 Algorithms performances

Contours detection method has a principal problem linked to shape extraction. Three main drawbacks of Canny contour detection and reconstruction were cited in the Paragraph 5.1 and dealt with: near object contour connection, bad contour closing and connection of object with possible contours artifacts.

The first problem causes a recall reduction value since near objects that could be kept are discarded and increases false negatives values. Contour bad closing is linked to partially detected contours problem that may be due both to bad Canny detection parameters or a bad illumination (for example in the camera edges). Some tests have been done to demonstrate that illumination in edges decreases and is not enough to highlight contours far from the center. Furthermore, around the images corners, far from the image centre, perspective projection acts deforming segmented objects.

If a lucid conveyor is used, local light reflections can appear in Intel camera RGB images. These bright reflections on dark backgrounds may generate high contrast small areas that map into contours artifacts. Contour connection procedure may connect these artifacts with objects compromising their blob shape and consequently their barycenter and orientation estimates. In

conclusion this method doesn't grant solidity in terms of reliable complete contours creation. It requires high contrast between background and objects and a conveyor belt possibly opaque. It also leads to bad performances since only isolated objects may be analyzed and this causes the recall parameter to drop remarkably.

The Background Subtraction method cannot be used with RealSense camera since depth maps are imprecise and not perfectly aligned with the RGB images. It's not possible to get a reliable segmentation of objects by applying a subtraction between a depth frame with objects and the background one.

The only case in which it works is with Basler camera. Results are promising, without shape matching procedure the algorithm is able to detect blobs of free isolated objects and to derive brycenter, orientation and verse information. On the other hand, it's not possible to analyze overlapping objects relations and this causes the efficiency of the algorithm to be significantly low.

Finally if the products are moving, like in current application, this method cannot be used with V section shape conveyor belts since the background model changes in time. It's not possible to resample every time the background in order to subtract it from the foreground depth images since the conveyor moves continuously.

Hence, this method cannot be used with Intel camera, it's able to identify only isolated objects and cannot be used with not plane conveyor belts. For these reasons also this method is folded in search of stronger ones.

Derivative Method

Some considerations have also to be done about algorithm strengths linked to shape-based matching performances. Derivative method can be applied also when not all objects inside the image are found. It could happen that hardly occluded objects are difficult to be identified or it's required to keep high threshold in the matching score during the research due to fact shape-based algorithm fatigue to orient right the shape. In these cases not all the objects are found along the research but only well oriented ones.

However, the derivative based algorithm is able to establish if the object is free or not. This property is very powerful since doesn't require the hard assumption of finding all the image objects during search. In fact objects' longitudinal depth pixel lines are extracted and analyzed one by one; information coming from other objects' shape positions does not influence the singular object pick up decision. For example if only the current object shape is identified (Figure 6.4) then only its shape position and depth image are required to establish if it's free or not.

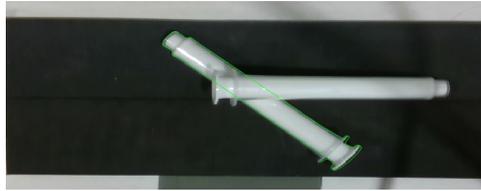


Figure 6.4: RGB Image with green contour in correspondence of detected object



Figure 6.5: 2D Object Depth map (blue) and its derivative (red)

Since the derivative maximum exceeds the threshold value then the object is discarded. High derivative magnitude means some other object is laying on the current one as is possible to see by Figure 6.5 .

The same can be said if, for example, only the upper one object would be found by the shape matching research (Figure 6.6). The algorithm is able to establish that it's free without information about where other objects are. In fact, looking at Figure 6.7, no derivatives peaks over the threshold imposed are present and the object is classified as free.

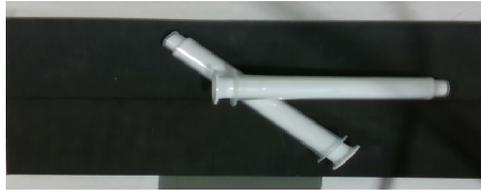


Figure 6.6: RGB Image with green contour in correspondence of detected object

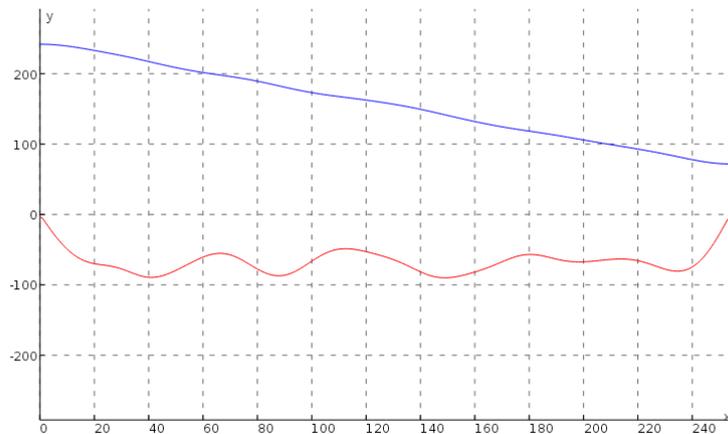


Figure 6.7: 2D Object Depth map (blue) and its derivative (red)

Common Area Method

The same things told in previous section can't be said for the Common Area method that relies on the strong assumption to find all the objects in the image. This is then a necessary condition both for simple identification of free objects and for having picking order for objects. It's a very strong assumption since it's difficult to always grant to find all objects in the images especially when hard occlusion cases happen.

Considering, for example, the unfortunate case in which only the highlighted objects are found, the pick order would result as the one shown in Figure 6.9 .

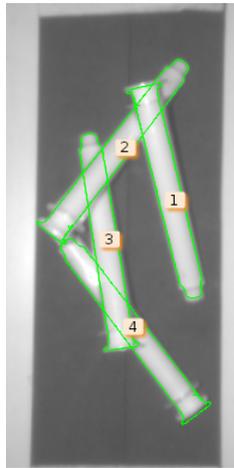


Figure 6.8: Ground Truth Ordering (threshold = 0.76)

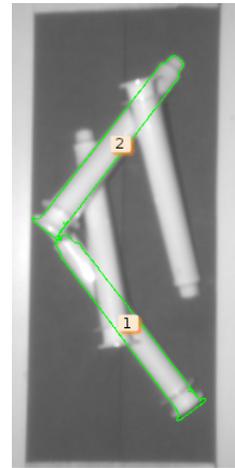


Figure 6.9: Wrong Ordering (threshold = 0.9)

Respect to ground truth image in Figure 6.8, the order shown in Figure 6.9 is different but also wrong, this would cause to select as free object a false positive object. As said in the problem specifications false positives are not accepted and so that situation does not have to happen. Low numbers of results can be attributed to high threshold parameters in the search algorithm since sometimes could happen that, with lower thresholds, shapes are found in areas in which there aren't objects. This is because contrast in the image may trick the research procedure as described in Paragraph 6.1 .

In this case comes in help also the possibility to filter the search area using depth image information as introduced in Paragraph 3.1 and shown in Figure 6.10 . The procedure consisted in reducing the domain in which the model is searched by the shape matching algorithm in order to reduce both the search area and the computation time. It's based on the detection of areas in which, for certain, no objects are present. By having the search filtered area is then very difficult to find false positives during the search. In this way, the threshold for the research can be kept lower, increasing the probability to find always all the objects but keeping low the probability of finding objects where there aren't.

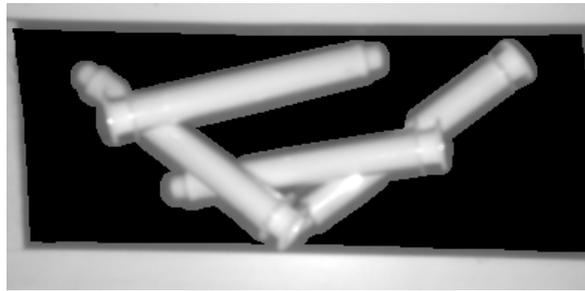


Figure 6.10: Grayscale filtered Image

MLP Method - Algorithm generalization

The MLP method relies on the training of a MLP network able to classify depth images based on the pixel color values. In particular this classification can be done thanks to the fact that objects always lie at the same distance with respect to the camera. The classification turns out to be useful since the depth image is classified in three areas: Background, Low and High. The Background labeled areas belong to the conveyor belt free parts and can be used to know where, for certain, objects have not to be searched. The Low labeled objects lie on the conveyor belt while the High ones are parts of objects laying on other objects. The classification performed and needed is approximative and not precise but anyway useful to know if objects are free or occluded. In particular free isolated objects are detected if they don't contain High labelled part, while free objects inclined because on top of another one are detected by looking at the High labelled area that they contain.

The method has a big drawback connected to the maximum number of objects interacting when an occlusion happens. In particular, at maximum groups of two objects may be examined by this algorithm. More details are given in Paragraph 5.6.2 .

In the unfortunate case the identification task have to contemplate also more than two objects interacting, two ways can be followed:

- The blob containing more than two objects is detected looking at blobs area and subtracted from the search area. In this way none of this object will be found by the search algorithm. This reduces the algorithm efficiency since false negatives are generated.

The algorithm works no matter if all objects are found during search. An estimation of three objects' minimum area has to be calculated for each different object for which this algorithm is adopted.

- An improvement of the algorithm is introduced to cover also the case of three interacting objects making more flexible the algorithm. The

idea behind the core algorithm remain exactly the same: the possibility to pick up an object comes from a threshold on the goldenrod area inside found objects contours. This improvement comes not for free since could happen that some false positive arises in three objects interactions. It can be seen in the example represented by Figures 6.11 to 6.14. In this case the object in Figure 6.12 is classified as occluded but the two objects highlighted in Figures 6.13 and 6.14 results grabbable since their areas are major of the threshold imposed in this algorithm. Indeed, the object of Figure 6.14 is wrongly detected as grabbable since it's occluded. In other words, an occlusion occurs among two objects that the algorithm classified as free.

The algorithm is able to detect this error if and only if all the three objects are detected by shape-based research. In fact, if two objects that should be picked up touch themselves, they are automatically discarded since it's not possible to establish which is higher respect to the other. It cannot happen, in fact, that the areas of two objects that have to be picked up have an interception area different from zero. In case only one of them is detected, like in Figure 6.14, a false positive would be generated and this is not acceptable.

Finally, interactions of four or more objects will be discarded in the way described in the previous case so by excluding the region in which they lie from the search area.

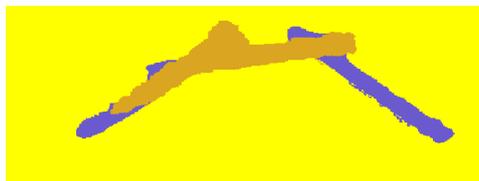


Figure 6.11: Classified Depth Image; Background (yellow), Low (violet) and High (goldenrod)



Figure 6.12: Occluded object contour with High classified area inside of it (yellow)



Figure 6.13: Free object contour with High classified area inside of it (yellow)



Figure 6.14: Occluded object contour with High classified area inside of it (yellow)

In order to calculate if objects are parallel or oriented respect to the conveyor belt, a simple solution was provided in Paragraph 5.2 . In that case it was easy to understand if an object was oriented or not since with at most two objects was involved in an overlapping relation and the High region area enclosed in the contour provides intrinsically that information. In particular if the High area is small the object is not oriented, if it's large the object is oriented.

Now, a not oriented object could be also on some other objects, having large High area enclosed in itself (for example in Figure 6.20). A solution that could be adopted in all the cases is to consider the standard deviation of the depth values inside the considered object: if it's low the object it's horizontal so not oriented, if it's high the object is oriented. Occluded objects have high standard deviation even if they are horizontal but this does not affect our algorithm since occluded objects have to be always discarded for the requests asked in current application.

Another simpler way for the orientation procedure, is to establish a threshold for the High enclosed area. In particular, considering grabbable objects:

- in case low High area is enclosed in the object then it's horizontal.
- in case very high area is enclosed then the object lies horizontally on top of others
- in other cases it's oriented

6.4 Static Results

With Intel Realsense camera, the Common area method cannot be used since the depth mapping result imprecise respect to the RGB image. Depth-RGB images alignment coming from the libRealSense library is not responsible of this bad alignment. Approximative depth mapping especially around objects contours comes from the low reliability of depth estimated. The reason is associated to High Density preset that lowers too much the confidence in matching estimates. Hence, the Filling Holes post processing filter carries the Filling Rate to 100 % and is very light weight but the filling quality results bad and without any basis on depth calculus.

By not choosing the High Density preset, for example using Medium Density or High Accuracy ones, the estimation quality of depth pixels increases but lowers the Filling rate, causing very large holes in the depth images. The bigger the holes in depth images the more is difficult to fill them in post processing phase.

In conclusion the best approach is the one using the High Density preset that, jointly with the Holes Filling procedure, grants overall satisfying estimates with 100 % Filling Rate. The Holes filling type used is the number 3: hole pixel value assumes the smallest among the valid five upper left and down pixel values.

Respect to the normal High Density preset, the "DS Second Peak Threshold" has been additionally lowered. By doing some tests, this change in parameters causes the Filling Rate to be increased. More details about this parameter, presets and post processing filters can be found in Paragraph 1.2.

6.4.1 Intel RealSense D435

Some significant objects have been chosen to test the effectiveness of the algorithms designed. From now on the results in static conditions using the Intel Realsense camera will be presented. No cases using the Common area method will be shown since, as said several times, the depth images coming from the Intel RealSense camera are not suitable for its usage. The main problem is connected to the fact that depth maps are imprecise and approximative around objects contours.

It's important to underline that tests results presented don't last only for examples presented in the following part. Even if precise statistics are not present about the working of the methods and recall rates, tests have been done for over 30 images of each product demonstrating the effectiveness of algorithms. As required the 100 % precision rate has been satisfied. Detailed and deep statistics will be illustrated for some of these objects in dynamic conditions.

Orange Syringes

These syringes are easy to be matched by the shape matching algorithm since high contrast is present between orange foreground objects and black background.

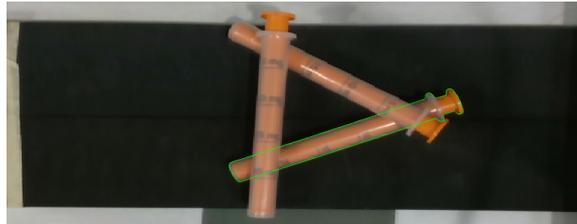


Figure 6.15: RGB Image

Derivative Method

In order to stress the Derivative algorithm the objects overlapping situation shown in Figure 6.15 is considered. The relative depth plots are shown in Figures 6.16 , 6.17 , 6.18. Overlapping situations are visible from humps and derivative peaks in Figures 6.17 and 6.18. Setting the derivatives peak threshold to 180, these objects are all detected as not grabbable.

On the other hand looking at the Figure 6.18, representing the occluded objects depth plot, it's evident both from depth map and from derivative plots that two objects are occluding the one considered. Of course this object will be discarded since the peaks of the derivative are in magnitude higher than 180. In the derivative plot 4 peaks are present, 2 positives and 2 negatives. For each hump in the depth map there are one positive and one negative peak in the derivative plot. This due to fact the hump has an ascendant and descendant phase. It's the case of the highlighted object in figure 6.15 that, in fact, it's occluded by two objects.

Looking at the Figure 6.16, no part of the derivative function assumes, in magnitude, values major than 180. The depth map is very homogenous and almost horizontal. As the near to zero derivatives values shows this product is not inclined. This plot belongs at the free horizontal object on the top of the other two.



Figure 6.16: Free Object Depth map (blue) and its derivative (red)



Figure 6.17: Not Highlighted Object Depth map (blue) and its derivative (red)

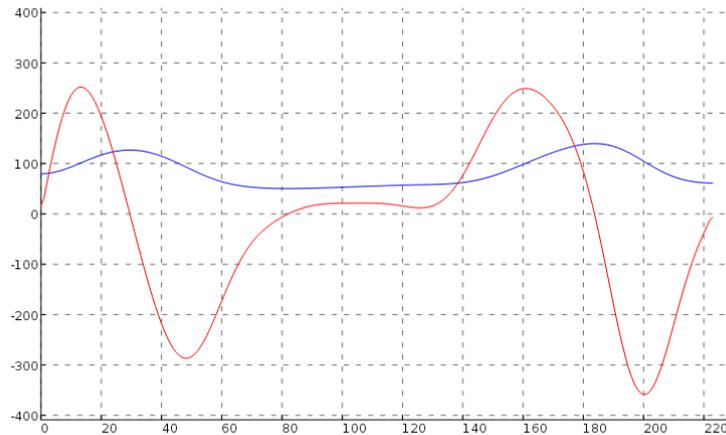


Figure 6.18: Highlighted Occluded Object Depth map (blue) and its derivative (red)

Remember that the threshold imposed is always compared with the absolute value of the derivative function. Both positive and negative peaks, in fact, are indexes of abrupt depth variations and so of the presence of one occluding objects. If the derivatives peak is only one then it means that the overlapping happens in an object extreme.

In general, in Derivative method, a necessary and sufficient condition to detect an overlap is to have at least one derivative peak over the threshold imposed.

MLP Method

The same configuration examined with Derivative method has been analyzed with the MLP one. In Figures 6.19 , 6.20 and 6.21 are represented the High areas interceptions with each detected object. Both objects in Figures 6.19 and 6.20 are detected as grabbable since the yellow area is major than the threshold imposed. Instead, the syringe highlighted in Figure 6.21 is discarded. Since the intersection between the two grabbable objects is different from zero an identification error is found. In fact it's not possible that two classified free objects have an overlapping part. For this reason all the objects inside the current blob are discarded.

The example just proposed is a case in which the Derivative method gives good results detecting the free and occluded objects. Instead the MLP method isn't able to manage right the detection of free objects giving wrong results. Anyway, since all the objects in the image have been found, the error is detected and all the objects are discarded.

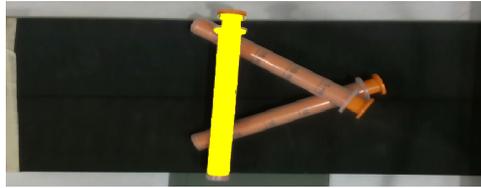


Figure 6.19: Occluded object contour with High classified area inside of it



Figure 6.20: Free object contour with High classified area inside of it

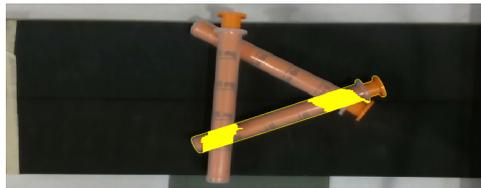


Figure 6.21: Occluded object contour with High classified area inside of it

An example of MLP method working with three objects is shown in Figures 6.22 , 6.23 and 6.24 . The occluded object presents a lower interception area with respect to the free ones that are almost completely filled by the High region. The two free inclined objects don't have a common overlapping area and so they can be picked up.



Figure 6.22: Free object contour with High classified area inside of it



Figure 6.23: Free object contour with High classified area inside of it

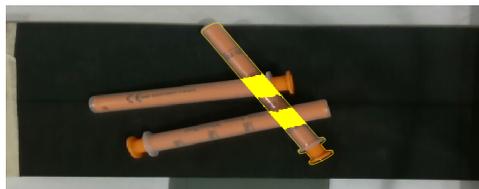


Figure 6.24: Occluded object contour with High classified area inside of it

Remember that the MLP method, if assuming to work with at most two interacting objects, doesn't require to find all the objects in the image to do a reliable complexive estimation. For "reliable estimation" is intended an estimation that doesn't provide false positives that is not an acceptable result in current application. Assuming to have more than three objects, like in the configurations considered in previous examples, the strict requirement of detecting all objects is necessary to have 100% precision.

The Derivative method shows to be very performant even in cases in which not all objects are detected. An object can be classified for certainty as free or occluded even if it's the only one found among all.

White Untextured Syringe

Derivative Method

This is one of the thickest syringes available among the one examined. What is expected, observing plots coming from the different algorithms, is to have most evident and clear results in this case.

White on black background images makes very easy the research and identification process. Also the orientation of the object is easy since the top of the syringe contour is different from the bottom one. The case analyzed is the one represented in Figure 5.37 that was used to present the results obtained with the MLP method.

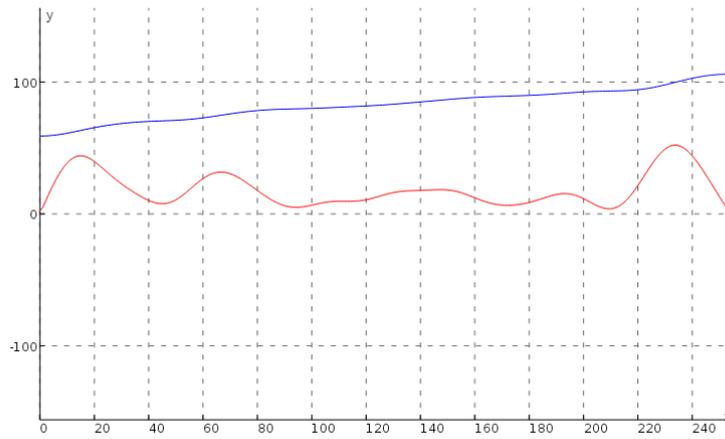


Figure 6.25: Free Horizontal Object Depth map (blue) and its derivative (red)

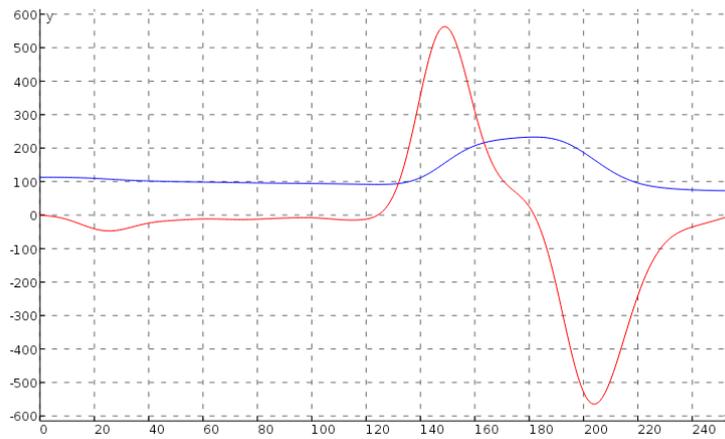


Figure 6.26: Occluded Object Depth map (blue) and its derivative (red)

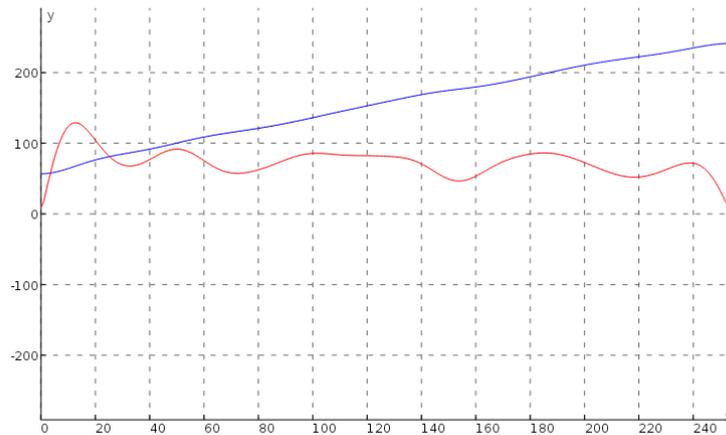


Figure 6.27: Free Inclined Object Depth map (blue) and its derivative (red)

In Figure 6.26 the 2D depth map of an occluded objects is shown, the derivatives peaks are very high in magnitude.

In Figures 6.25 and 6.27 are presented depth and derivative maps of the two free objects. In this case the threshold to detect occluded objects have been set to 250. Peaks of the derivatives are sharply lower respect to the threshold imposed and it's easy to classify them as free. Gaussian filters used to smooth the function acts very well, smoothing the depth function enough to grant no local discontinuities but preserving global oscillations. From depth map in Figure 6.27 it's evident that the object is inclined while the one of Figure 6.25 is not.

MLP Method

MLP method is applied also to a similar configuration of the one before in which one free object is horizontal and one is oriented. The results are shown in Figures 6.28 and 6.29 . The left free syringe is classified as free since it doesn't contain goldenrod areas.

The free inclined object and the occluded one are easy to be distinguished since they contain very different quantities of High areas. In this case it's easy to establish a good threshold for the area magnitude to distinguish free from occluded objects.

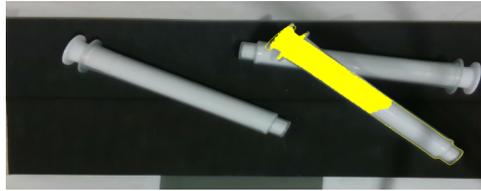


Figure 6.28: Occluded object contour with High classified area inside of it

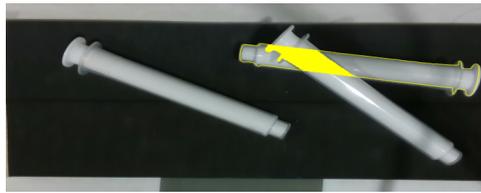


Figure 6.29: Free object contour with High classified area inside of it

Small Syringe

Some objects, among all the available ones, have been chosen to test algorithms and stress them as much as possible. For example, one of the thinnest syringes available has been considered. Its thickness is around half centimeter and its length is about 8 cm.

Problems that can arise are mainly connected to the fact that depth mapping error becomes more and more compatible, in terms of magnitude, with the thickness of these objects and may compromise results. This means that height variations due to disturbances may change significantly the object depth mapping creating high and not real discontinuities.

Talking about derivatives, the variation in depth in overlapping cases could no more result as evident as with big syringes. At the same time, MLP networks may fatigue to classify depth maps due to local pixel oscillations and uncertainties. As remarked several times, High Density filter is used jointly with Holes Filling post processing filter. Since the High Density filter increases Filling Rate by lowering the matching votes threshold, more uncertain depth mapping arises around the contours. And since objects thickness is low, the major part of the objects are mapped in depth with low reliability.

Derivative Method

First of all some plots regarding the Derivative method are presented to show and comment results.



Figure 6.30: RGB Image

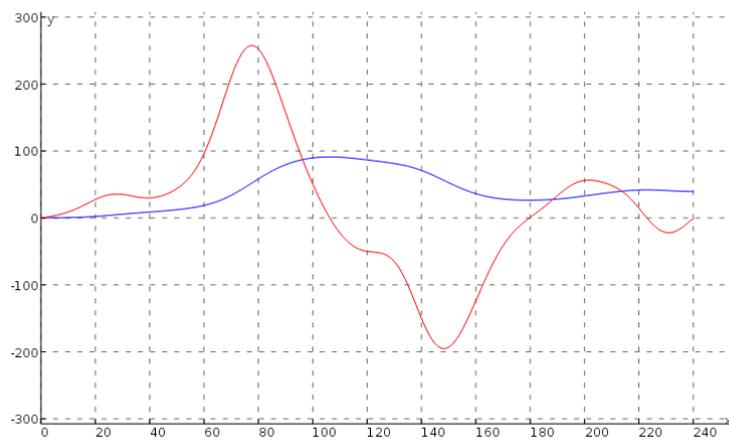


Figure 6.31: Occluded Horizontal Object Depth map (blue) and its derivative (red)

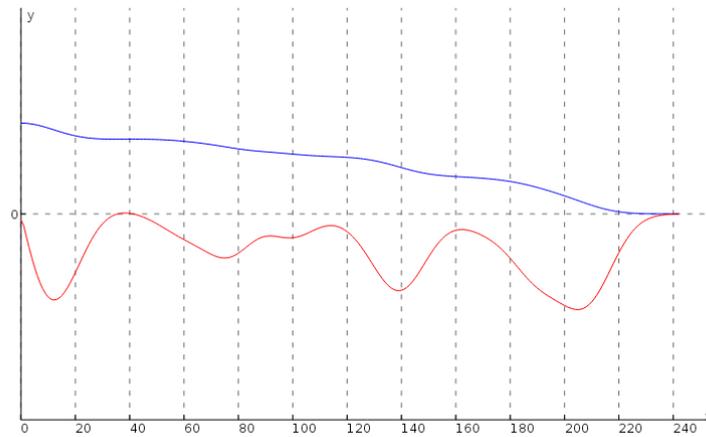


Figure 6.32: Free Horizontal Object Depth map (blue) and its derivative (red)

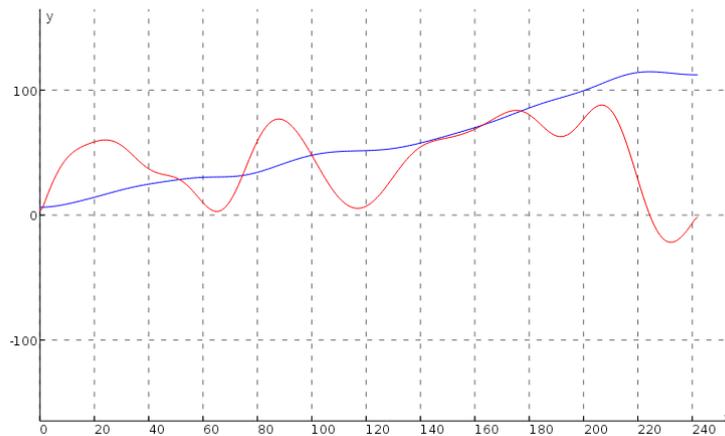


Figure 6.33: Free Oriented Object Depth map (blue) and its derivative (red)

From Figure 6.32 it can be seen that the derivative plot in case of a free object is low in magnitude and the depth map is still well estimated by Realsense camera. The same can be said for the free inclined object in Figure 6.33, where it's evident the object inclination observing the increasing of the 2D depth function and the almost always positive derivative function. At maximum the derivative function assumes values less than 100.

By considering the occluded object plots in Figure 6.31, it can be noticed that the derivative peaks have values over 180. The depth map shows

the presence of one object passing on it and in fact the derivatives assume in magnitude two high and clear peaks. 100 and 180 are far from each other and a threshold can be set for example at 120 to distinguish occluded cases by oriented ones.

Intel RealSense depth maps for Derivative method result still reliable and strong, giving high detailed plots even considering objects with low thickness. The method is flexible also for cases of small objects.

MLP Method

In Figure 6.34 is shown the classification output of configuration in Figure 6.30 .

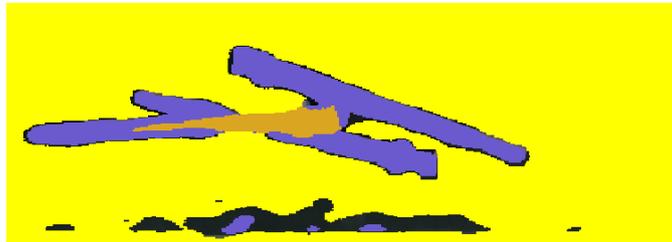


Figure 6.34: Classified Depth Image



Figure 6.35: Occluded object contour with High classified area inside of it



Figure 6.36: Free object contour with High classified area inside of it

Differently from other classification maps, the quantity of black pixels

is high. Black pixels are pixels unlabeled since the neural network is not trained to recognise them. This phenomenon can be attributed to the fact during the training phase, when the samples are chosen for each label, some depth colors haven't been included. Then these colors, when analyzed by the network, cannot be labelled. Anyway it's not a problem since the important information dealing with objects depth color labelling are present and so the method can work.

Of course the free object has only violet labels and so it's classified as free. The occluded one (Figure 6.35) and the free oriented one (Figure 6.36) have more or less the same High pixels areas inside of them. By doing some tests it comes out that using these areas to establish a threshold is not very reliable since areas are not evidently different and could happen that occluded objects are classified as free (false positives).

To avoid having false positives the threshold chosen have to be small and this causes to have a lot of false negatives. As a consequence, high false negatives number lead to low recall values.

An improvement adopted in these cases is to consider a thicker region with respect to the contour one to calculate the interception with the High region. In this way can be demonstrated that the two areas obtained are more different and a safer threshold can be set.

In general, for this kind of objects that are thick the MLP method is weaker with respect to the Derivative one. The area threshold is difficult to be set and it's not granted to have 100% precision rate. The Derivative method is very strong and reliable giving with static objects a Recall reateo of 100% on 94 objects in 30 images.

In addition, in case of Figure 6.37, the objects are all analyzed by the Derivative method while, MLP method would have discarded them since more than 3 objects are present. In this case the Derivative method is able to detect the two oriented free objects.

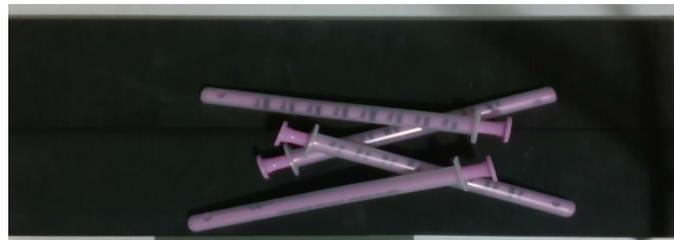


Figure 6.37: RGB Image

From Figure 6.35 can be seen that the High mapped area is not coincident with where the object effectively is. This is because of the approximative depth mapping described in the introduction of this Paragraph.

Big head syringe

These are particular kind of syringes that has not a perfectly cylindrical shape but has an additional bigger head part. In Figure 6.38 an example of some of these product is shown. From this image can also be noticed that perspective deformations change evidently their shape and so the contours of the objects. In order to solve and compensate for these deformations, the deformation and min/max scale parameters of the shape matching algorithm can be opportunely tuned. The research time can be slightly increased but the quality of the matching increases significantly.

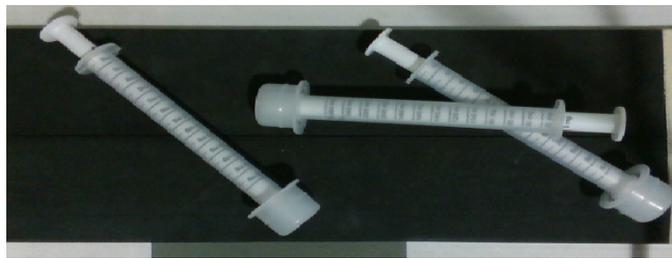


Figure 6.38: RGB Image

Derivative Method

The problem that could appear trying to use algorithms with these products is connected to the fact that the depth mapping varies a lot along the object major axis. Talking about Derivative method, what it's expected is to have a high derivative peak in free objects caused by the abrupt passage between the depth of the body and the depth of the head.

An example of this phenomenon is shown in Figure 6.39 where is present the depth and derivative plot of the free left object. Is possible to notice the depth discontinuity on the right extreme of the depth plot (blue) that reflects the section shape of the product. As expected the derivative assumes a peak in correspondence of the depth variation.

By examining Figure 6.40 it's possible to see the depth plot of the occluded object. In correspondence of the depth hump, two peaks of the derivatives arise. The magnitude of these peaks is about 3 times the peak value observed in the free object plot. This means that, even in case of having depth discontinuities in the depth map of a free object, a threshold to distinguish free and occluded objects can be found. In particular the threshold has been set to 250.

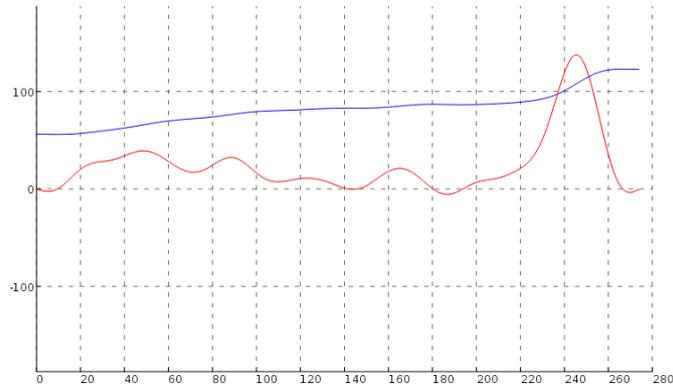


Figure 6.39: Free Horizontal Object Depth map (blue) and its derivative (red)

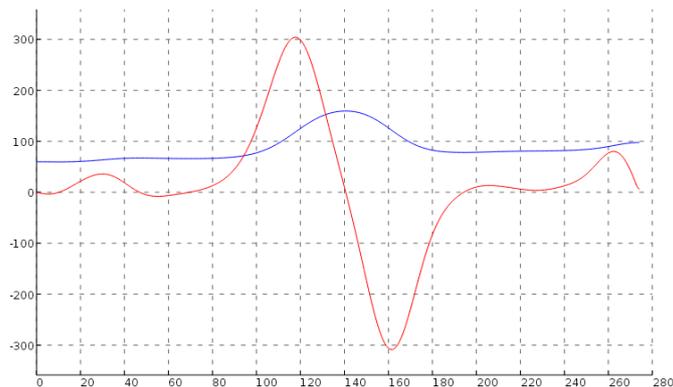


Figure 6.40: Ocluded Object Depth map (blue) and its derivative (red)

MLP Method

The same configuration can be examined with the MLP method. As can be seen in Figure 6.42, free objects show the head classified as High. This result was expected since the upper image parts are at the same height of the syringe head. Anyway, looking at the overlapping object (Figure 6.44), it's almost entirely labeled as High respect to free and occluded object that has a lower High area enclosed in its contour (Figure 6.43). The enclosed areas between occluded and free oriented objects are very different and hence a strong threshold can be established.

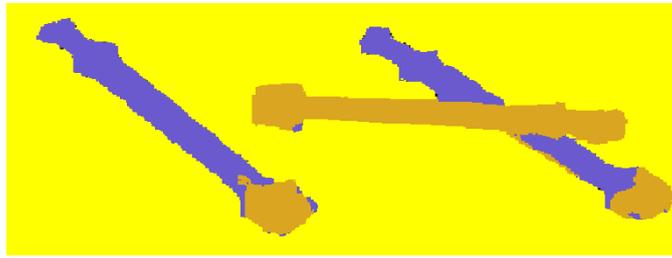


Figure 6.41: Classified Depth Image

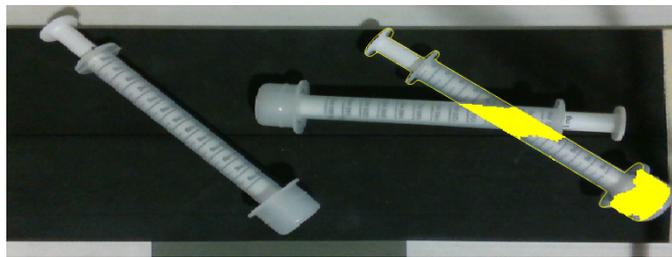


Figure 6.42: Occluded object contour with High classified area inside of it

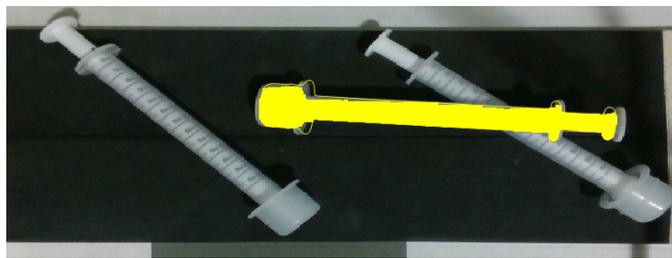


Figure 6.43: Free object contour with High classified area inside of it

For the MLP method, in case of more than two interacting objects, the techniques used to establish if an object is oriented or not was based on the depth standard deviation or on enclosed High area. In this case the standard deviation data is no more reliable and two more efficient methods have been studied.

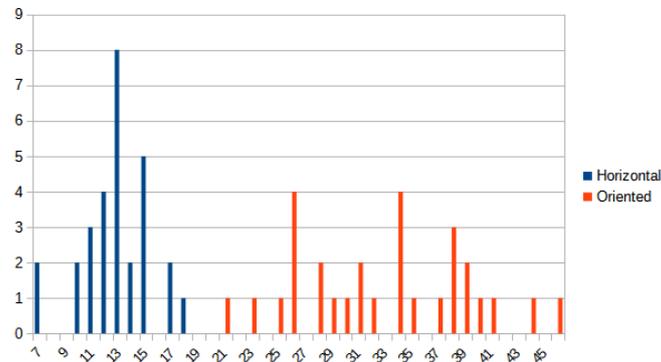


Figure 6.44: Objects depth standard deviation values

In the Histogram in Figure 6.44 are represented standard deviation values collected from a bunch of inclined and plain syringes of the type now considered. It's very difficult to establish a strong (with high margins) threshold value able to classify oriented objects with respect to not oriented ones. The cause of the high dispersion of standard deviations data are due to: local oscillations of datas, imprecision in the depth estimates around contours and the not uniform depth maps of these objects.

A solution to counteract bad depth estimation of contours is to calculate the standard deviation in a shrunk area respect to the model one. In this way depth values around contours are not considered. In addition, derivative information coming from the mean value inside the shrunk object area can help in establishing object orientation.

Let's consider the histogram in Figure 6.45, showing the occurrence of mean depth values inside objects classified, based on their orientation. Without considering the last two red data a clear division between oriented and not oriented objects can be set in a stronger manner with respect to the standard deviation case.

Objects horizontal that have high means are represented by free objects lying on other objects (they are higher respect to the conveyor belt depth level). Doing several tests showed that, in these particular cases, the objects have very low standard deviation and so can be prefiltered as horizontal basing simply on their standard deviation value. In synthesis very low standard deviation means the objects are horizontal. Then, if their mean value is low, they are horizontal otherwise they are oriented.

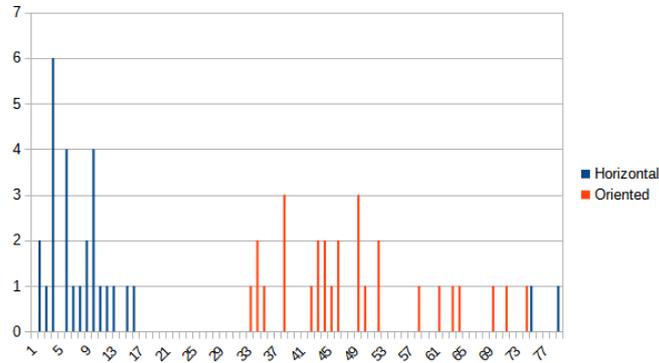


Figure 6.45: Objects depth mean values

In alternative, considerations about enclosed High areas and standard deviation can be adopted to detect oriented objects in a simpler way. The decision procedure can be divided in two parts: if an object has high standard deviation and its enclosed High area is large then the object considered is oriented; otherwise if the standard deviation is low (the object lies horizontally on others) or if the enclosed High area is small (the object lies on the conveyor belt) then the object is not oriented.

It's possible to conclude that Derivative and MLP approaches are flexible enough to be used also for this kind of objects, no matter if the depth map has higher variability with respect to previous ones. Anyway MLP, if complex objects interactions arise, grants lower recall values and, as a consequence, lower performances in terms of ratio between free picked up objects and total free objects. As shown in this chapter also the objects orientations becomes complex with more objects interacting.

Long Syringe

A final example is given considering another syringe type that is longer about two times the other ones. A consideration about camera distance has to be made in this case. As will be said also in the Dynamic tests section, the length of a syringe is about the width of the FOV of the camera. This fact could cause having low quantities of frames in which the objects are not effectively touching image border. These objects cannot be considered for further analysis. A solution to solve this problem could be to set more far the camera from the conveyor, in this way the FOV is increased and more frames are available for the detection process.

Apart from that, even for this syringe type, the algorithm performs well. One result dealing with Derivative method and one for the MLP will be provided to show another proof of the flexibility of this method. The threshold for the Derivative method has been set to 180.

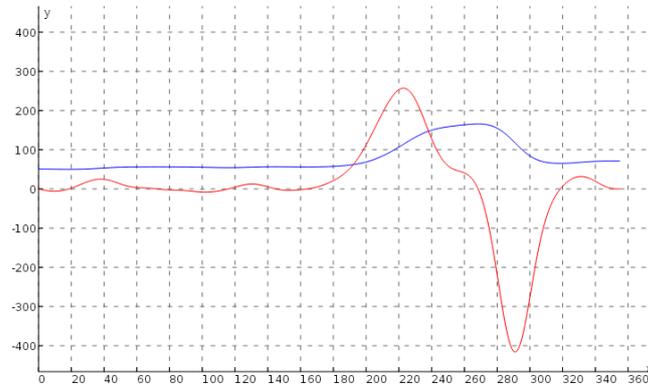


Figure 6.46: Occluded object Depth map (blue) and its derivative (red)



Figure 6.47: Free Horizontal object Depth map (blue) and its derivative (red)

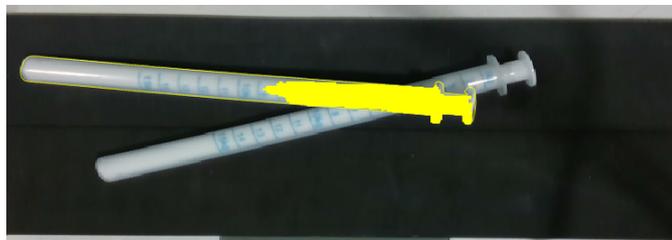


Figure 6.48: Free object contour with High classified area inside of it

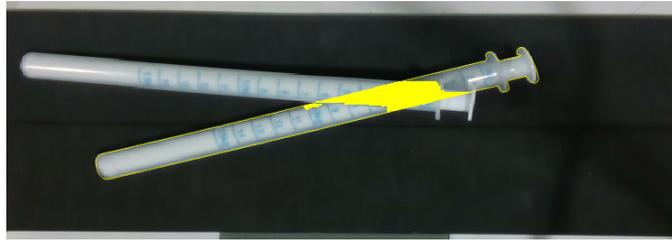


Figure 6.49: Occluded object contour with High classified area inside of it

For static objects both the MLP and Derivative algorithms show satisfying results. In the next Paragraph results with Basler camera will be illustrated trying to enhance eventual advantages, limits or weaknesses of the methods.

6.4.2 Basler blaze 101

A significant object has been chosen to put in evidence main problems given by this camera. Holes in images caused by local saturations compromises the depth images used for the designed algorithms. As will be shown in the case of the white untextured object, already used in Intel camera tests, algorithms proposed can be still used but are more prone to errors and not very reliable. Remember that this object is the one giving best results with algorithms. Having troubles in the application of algorithms with it means that these troubles will be much more present also with other objects.

A solution has been designed to solve holes problem, it will be shown in the final part of this Paragraph and will be used and tested in dynamic conditions.

White untextured syringes

The problem of holes that can arise in the depth maps and generate false negatives have to be solved. Their presence lowers the algorithms' efficiency but doesn't compromise their application. Some testing about different methodologies of filling holes have been done being aware also on the time spent to do computations. Filling procedures from the most elementary to the most complex ones have been tested. Among them the most efficient, also in terms of computation, are based on filling the black pixel with: maximum or minimum or mean or median depth value among the current pixel neighbors.

With Intel RealSense the pixel filling is based on the minimum depth value among the neighbor, in that case the filling necessary was approximative since the importance of that procedure was to have 100% Filling Rate

with not precise depth estimates.

On the contrary, now, the filling procedure is required to be as most precise and continuous as possible. In fact all the three analysis methods considered are based on the depth information in the central parts of objects.

The Derivative method uses depth pixel information on the major axis of the object. Considering the depth map along the axis of the object in Figure 6.50, the resultant depth plot with its relative derivative is represented in Figure 6.51 .

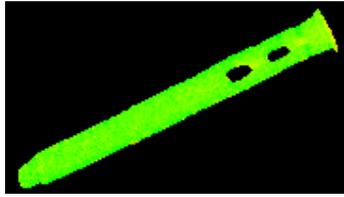


Figure 6.50: Color Depth Image

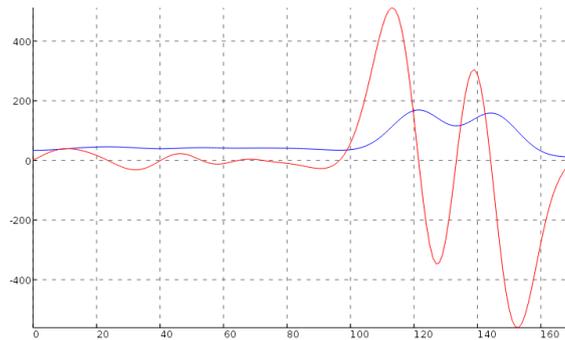


Figure 6.51: Occluded object Depth map (blue) and its derivative (red)

From the plotted derivative function it's evident the presence of high derivatives in correspondence of depth black holes. The derivative peaks are very high and this means that this object would be discarded. Holes in depth mapping causes, then, the increment of the false negatives.

Holes in images are generated since light hits objects in particular positions and causes the reflection of the beam directly back to the camera sensor. It's not always told that they appear in collected images. For example in Figure 6.52 it's represented the depth image in gray scale of two syringes. No holes are present on the objects and from the derivative plots in Figures 6.53 and 6.54 is clearly possible to detect respectively a free object from an occluded one.

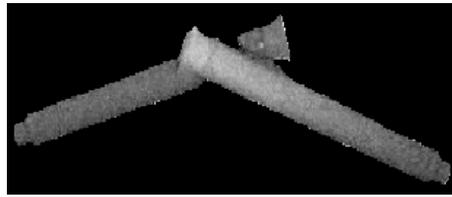


Figure 6.52: Gray scale Depth Image

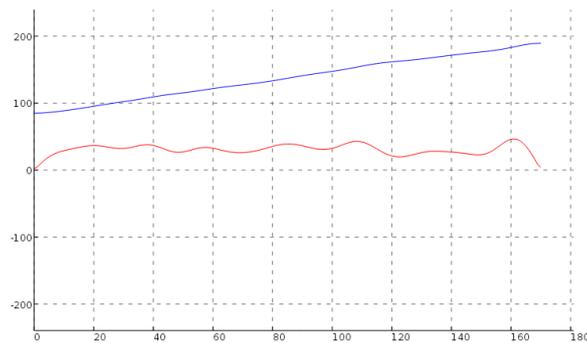


Figure 6.53: Free object Depth map (blue) and its derivative (red)

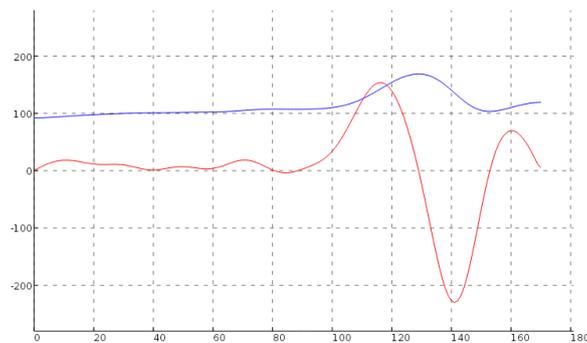


Figure 6.54: Occluded object Depth map (blue) and its derivative (red)

Talking about Common area method, the estimation of mean value along reoriented areas gives bad or wrong results if the common area includes holes. A weak solution adopted is to discard objects if too many black points are present in the Common area. This would increase the percentage of false negatives.

An example of a bad depth image in which objects are discarded instead of analyzed is represented in Figures 6.55 , 6.56 and 6.57 . In Figure 6.55 the

gray scale depth image of the objects is represented where, in green contour, is highlighted the common area that can cause troubles in the algorithm. In figure 6.56 is represented the objects picking order if the Common area algorithm is applied and objects are not discarded. The order generated is wrong and is caused by holes present in the common area.



Figure 6.55: Gray scale Depth Image

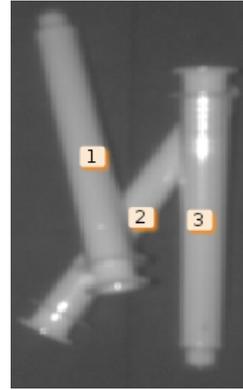


Figure 6.56: Wrong Ordering



Figure 6.57: Common Area

Finally, in the MLP method, the classification results would map holes as background labels causing mapping discontinuities in the classification maps. Some adaptations to this problem have been studied based on the detection and correction of the wrong labeled holes. In this way, the algorithm is able to work even with holes but it's very precarious. A depth image filling procedure may betterize results.

Hence, a filling procedure is required to increase the recall parameter and the consequent performances of algorithms. In order to know the areas to be filled, the confidence image is fundamental. First of all the confidence image is binary thresholded to eliminate low confidence values, then this image is dilated creating a binary mask.

The filling procedure is applied at black pixels in the grayscale depth image in correspondence of pixels that belongs to the mask just created. In this way the resultant object's depth map doesn't contain holes. The depth map around objects contours is wrongly estimated since the mask

is bigger than objects and causes background points to be included in the filling procedure. This is not a problem since the analysis done deals only with depth estimates inside the objects.

Among all the methods to fill presented the best one is the one based on the median values of pixel neighbors. The method consists in keeping the 8 pixels values around the pixel considered. The median value among these values is calculated (without considering 0 valued pixels) and assigned to the black pixel. This procedure takes few milliseconds that is a small amount of time with respect to the overall algorithm timings. Other filling procedures only reduce but not solve the problem caused by holes in depth maps.

The derivative method applied in one image corrected with the holes filling procedure just described is shown in Figures 6.58 and 6.59 .

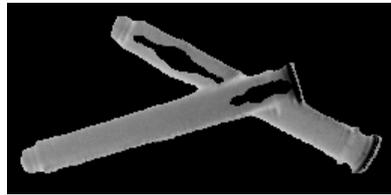


Figure 6.58: Depth Image with holes



Figure 6.59: Depth Image with filled holes

6.5 Dynamic Results

After doing experiments in a static framework, where products were positioned on a standstill conveyor belt, some tests have been done in order to evaluate algorithm and camera performances with moving objects. The velocity trajectory used to command the motor follows a trapezoidal waveform generating an alternated motion. The conveyor starts accelerating up to a certain velocity, then its velocity is kept constant and finally starts decreasing until the conveyor stops moving. The top value of velocity reached by the conveyor belt used for tests is about 11.6 m/min .

Spatial and Temporal filters couldn't be used with Intel Realsense camera because they cause wrong depth estimation when objects are moving. Images obtained without these filters result more affected by disturbances and oscillations of estimated parameters.

Basler camera comes with a particular Temporal filter able to detect motion and change only depth values that don't vary sharply in time. This filter could be used in the following examples depending, case by case, on the advantages it carries. In general, it doesn't compromise the Common Area method results even if could partially distort depth estimates. With MLP method this filter is not suggested to be used since would worsen the labelling results.

One way to increase depth estimates precision, when Temporal filter is enabled, could be also to enable Fast Mode. In this mode, the camera is able to provide up to 30 FPS with a lower precision and with more oscillating results. On the other hand, more frames are available for the Temporal filtering in the same span of time giving more overall precision on estimates.

In conclusion, with Basler camera, Temporal filter jointly with Spatial filter and Fast Mode could give advantages with designed advanced algorithms but it depends case by case.

An important requirement of this identification task is that, in each sequence, have to be present at least one image containing objects that are not crossing the borders of the camera field of view. It has been shown that the distance chosen (30 cm) from the conveyor belt is far enough to grant having this assumption respected for the major part of objects. Otherwise the camera distance has to be increased for those objects that sometimes are too long or too big to be impossible to be detected at least once. Given this assumption tests can be run being aware that miss detection of objects have not to be attributed to the fact that the camera collected too few frames to find them.

The mask created to detect objects partially included in images allows to discard objects crossing them and avoid searching objects far from the center of the image that are more difficult to be matched since they are more influenced by perspective distortion. Remember that this mask is strictly necessary since it's very difficult to match objects crossing the image borders especially when distorted by perspective. In addition, the overlapping relations cannot be determined when depth maps are partially available or it's not completely clear the number of objects interacting with it. The only treated cases are the safe ones, when objects examined are all included in the camera field of view.

6.5.1 Intel RealSense D435

Auto exposure doesn't keep into account possible motion blur effects but optimizes parameters in order to get good illumination, modifying only the exposure value. The exposure time in the region of interest is regulated based on the "Mean Intensity set point parameter". Also the ROI area is customizable in order to set the area in which objects will lie. In order to get sharp images even when objects are moving on the conveyor belt, it's necessary to disable the auto exposure feature and set parameters manually.

By reducing exposure time to avoid motion blur in images, the RGB image becomes of course darker since less light is caught by the camera sensor. To compensate this effect gain and brightness can be increased in order to enlighten the image. Increasing gain means increasing noise in image and this can compromise the image quality especially when shape matching detection has to be done. The exposure time has to be low depending on how fast the object goes on the conveyor. In the case considered, by setting $80 \mu s$ as exposure time and compensating with gain, images provide good brightness quality, are sharp and have low noise levels.

Tests effectuated demonstrate that noise doesn't change shape matching algorithm performances and increases detection quality with sharper images. Since shape-based matching algorithm searches for features related to objects contours, a motion blurred image generates bad matching results. In fact, it's not possible to precisely locate the model in the image and object localization is very uncertain.

Intel Realsense camera has two separated modules to provide RGB and depth images with independent and dedicated parameters settings. Hence, both RGB and depth modules have to be calibrated in order to avoid having motion blur problems. Since depth module bases depth estimates in stereo matching, the frames collected by left and right imagers could also suffer from motion blur problems. Anyway, no modifications in depth module parameters have been applied to the depth module obtaining good depth mapping even with moving objects.

In conclusion, the only parameters changed to compensate motion blur are related only to the RGB module and consists in Gain, Exposure time and eventually Brightness, Contrast and Sharpness.

White untextured Syringe

In Figure 6.60 an example of product discarded is shown. In fact, the product that is crossing the image on the right cannot be analyzed since it's not completely inside the image.



Figure 6.60: Object partially included in RGB Image

As described also in Paragraph 5.3.2 a blob containing an object that has a part in common with image borders is not considered in the final search area. It's done because of lack of precision in the shape detection procedure of the crossing object and lack of information to manage eventual interactions of this object with other ones not included in the image. The search image finally obtained is shown in Figure 6.61.



Figure 6.61: Filtered RGB Image

In this image the search area doesn't include the right object area and then it won't be found by the search algorithm. Since the filtering procedure is based on depth ranges there could be parts of the image, like in this case, that don't contain objects but carry depth values in that range. This is due to the fact, for example, the conveyor belt is not perfectly at the same distance with respect to the camera.

MLP Method

No evident differences in quality of results and depth classification are present respect to the correspondent Static case. The High area enclosed in occluded and free objects is evident and the threshold between areas can be easily established. The images collected are sharp enough to give precise identification results. In Figures 6.62 and 6.63 are presented MLP classification results for two moving objects: one occluded and one free.

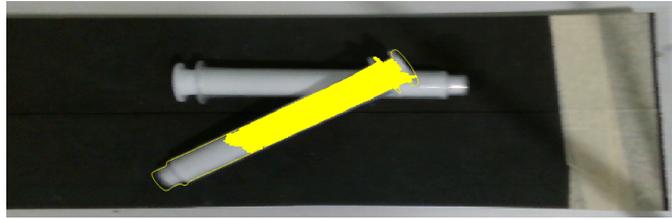


Figure 6.62: Free object contour with High classified area inside of it

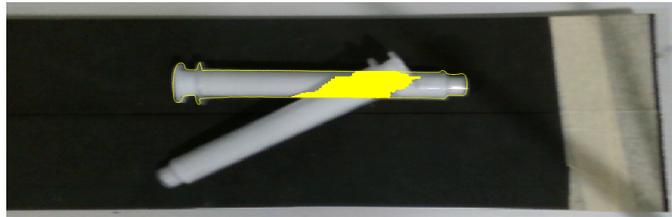


Figure 6.63: Occluded object contour with High classified area inside of it

Remember that the camera with MLP method has to be set parallel to the conveyor belt, in this way the training procedure can work and classify a certain depth range with a certain label. Vice versa if this doesn't happen the method can't work and the classification would give wrong results.

Derivative Method

Also using the Derivative method with moving objects doesn't compromise estimates and results. In Figures 6.64 and 6.65 are shown respectively the depth and derivative plot of the free object and occluded one. No hard discontinuities are present in the free object depth map and the derivative stays under 100.

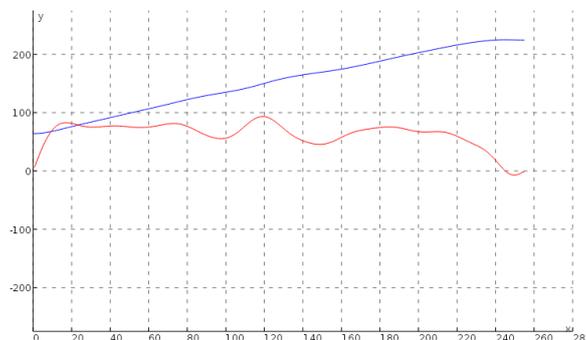


Figure 6.64: Free object Depth map (blue) and its derivative (red)



Figure 6.65: Free object Depth map (blue) and its derivative (red)

In the derivative image, instead, the two peaks reach values over 350 (in magnitude).

A sequence of several images have been collected and, setting the threshold to 250, the overall precision percentage obtained is 100 % and the recall 95 % .

Of course also the MLP method is set to have 100 % precision and gives 100 % precision rate. MLP and Derivative Method tests are done with the same test set composed of 213 images with 133 free objects. It's also important to underline the recall rate only for free objects since algorithms based on the usage of simple 2D cameras give 100 % recall rate for single objects but the management of overlappings becomes difficult.

False negative counts are not composed of objects not completely included in the images. Objects entering or exiting the camera FOV are not entirely described in depth and it can't be said for certain if they are free since their hidden part may present overlappings. In the Tables below results are resumed for this object in the two methods. Numbers between parentheses are referred only to free isolated objects.

	Graspable	Grasped	Not Graspable	Not Grasped
Derivative Method	133 (74)	127 (74)	59	65
MLP Method	133 (74)	133 (74)	59	59

Table 6.1: Derivative and MLP Methods results

	Precision	Recall	Isolated Objects Recall	Total Images
Derivative Method	100 %	95 %	100 %	213
MLP Method	100 %	100 %	100 %	213

Table 6.2: Derivative and MLP Methods statistics

Since the depth mapping is more imprecise and difficult with moving objects more black and uncertain areas, especially around contours, are present before the filling holes procedure. Thanks to Filling procedure, a very approximate estimate of depth is given in these points. Anyway it could happen that these estimates are wrong. MLP method is not very affected by these errors since it relies on depth labeling results that can be not completely precise. Instead, the Derivative method can be affected by wrong mapping since it may cause high discontinuities in the top and bottom parts of objects. In case of a free object, this bad mapping would generate high derivatives and, as a consequence, false positives cases.

This is the main explanation of the arising of some false positives in the Derivative method. As can be seen by Table 6.2, the Derivative method recall value is 95 % and the MLP method has 100 % recall value. MLP method is more immune to bad depth mapping caused by Filling procedure respect to Derivative one.

Orange Syringe

No particular problems arise in case of orange syringes, the detection is very easy since orange color is very distinguishable respect to black one and the model of this syringe can be extracted very well. Also with this object the methods that can be applied are the Derivative and the MLP ones.

An example of depth plots for free and occluded objects are represented respectively in Figures 6.66 and 6.67.

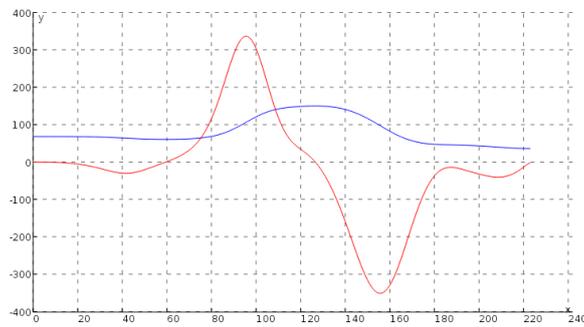


Figure 6.66: Occluded object Depth map (blue) and its derivative (red)

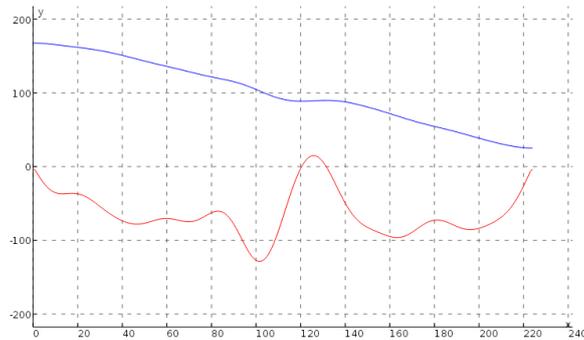


Figure 6.67: Free object Depth map (blue) and its derivative (red)

In Figures 6.68 and 6.69 are presented High interception areas derived from the MLP classification of the depth map. The two areas are evidently different in dimensions and a threshold can be easily determined.



Figure 6.68: Free object contour with High classified area inside of it



Figure 6.69: Occluded object contour with High classified area inside of it

Overall statistics obtained for this object are resumed in the Tables below. Numbers between parentheses are referred only to free isolated objects.

	Graspable	Grasped	Not Graspable	Not Grasped
Derivative Method	123 (91)	123 (91)	32	32
MLP Method	123 (91)	123 (91)	32	32

Table 6.3: Derivative and MLP Methods results

	Precision	Recall	Isolated Objects Recall	Total Images
Derivative Method	100 %	100 %	100 %	237
MLP Method	100 %	100 %	100 %	237

Table 6.4: Derivative and MLP Methods statistics

Long Syringe

As presented in the introduction of this Paragraph, the camera distance with this kind of long object plays a fundamental role. The more the camera is set far from 30 cm the more are the frames exploitable for depth algorithms analysis. At 30 cm happens that, for some objects configurations, no frames are available since objects are always crossing image borders. In addition, the more are the frames in which objects can be analyzed, the faster the conveyor belt can be driven. The increasing of the conveyor belt speed reduces the possibility of collecting images with objects completely inside of them.

Both the Derivative and the MLP method continues to work well even with these objects, the derivatives peaks are not as high as the ones of white untextured syringes but are anyway evident. In Figures 6.70 and 6.71 an example is shown.



Figure 6.70: Free object contour with High classified area inside of it



Figure 6.71: Occluded object contour with High classified area inside of it

No particular depth mapping problem arises. As can be seen in Figures 6.72 and 6.73 , even if objects are moving the depth mapping is not compromised.



Figure 6.72: Free object Depth map (blue) and its derivative (red)



Figure 6.73: Occluded object Depth map (blue) and its derivative (red)

Statistics and results are resumed in the Tables below. Numbers between parentheses are referred only to free isolated objects.

	Graspable	Grasped	Not Graspable	Not Grasped
Derivative Method	102 (71)	102 (71)	32	32
MLP Method	102 (71)	100 (71)	32	34

Table 6.5: Derivative and MLP Methods results

	Precision	Recall	Isolated Objects Recall	Total Images
Derivative Method	100 %	100 %	100 %	204
MLP Method	100 %	98 %	100 %	204

Table 6.6: Derivative and MLP Methods statistics

Timings

For objects shown until now with Intel RealSense, some qualitative results in terms of timings are estimated in Tables 6.7 and 6.8 . Remember that time measured is not the effective time needed to execute the algorithm in the robot packing machine. Time measurements are affected by processes

preemptions and memory accesses and so considerations can be done only in a qualitative and rounded up way. More details about timings measurements are reported in Paragraph 6.2 in the "Computational timing" section.

	Search Time	Computation time	Overall Time
White syringe	35,2	3,5	38,5
Orange syringe	6,8	2,5	9,3
Long syringe	319,6	3	322,6
Small syringe	11,5	2,5	14
Big head syringe	30,9	2,5	33,4

Table 6.7: Derivative Method Timings in ms

	Search Time	Classification Time	Computation time	Overall Time
White syringe	33,3	10,4	1,5	45,2
Orange syringe	5,4	16,2	7	28,6
Long syringe	323,1	10,0	1,5	334,6
Small syringe	17,7	10,9	2,5	31,1
Big head syringe	188,0	10,4	2	200,4

Table 6.8: MLP Methods Timings in ms

In the Computation column timings for the processing of one object are presented comprised of the time needed to prefilter out objects crossing image borders. If more objects are present in the image then timings expressed in that column have to be multiplied for the total number of objects analyzed. As said before these timings vary case by case and so only estimates to give time dimension orders can be done.

At the Total times shown the exposure time has to be added jointly with the Filling procedure time that needs about 3 milliseconds to be realized.

The main observation that can be done deals with the high magnitude difference between computation timing of the research procedure with respect to the rest of the computation time. The more evident case is the one of the long syringes. A detailed and fine design of the research procedure is needed to reduce at the minimum the impact of this procedure on the overall computation time.

The additional time required for the classification procedure is around 10-20 ms, that is an acceptable time considering its impact on the overall computation time. Of course the time expressed in this column deals strictly with the classification time, the training phase and the loading of the network are done in offline computations.

In conclusion Intel Realsense camera, if well calibrated in terms of exposure time, gives Depth and RGB images without motion blur effects. The timings required in using Derivative and MLP methods are expected to be similar also with the cases shown in Static conditions. As already said before, the Common Area method cannot be used with this camera since objects are not well depth defined around contours.

In Static tests objects configurations considered were complex and far from reality cases. This has been done to test strength and solidity of the algorithms designed. Object configurations considered for dynamic cases deal with configurations in which at most two objects interact in an overlapping situation. This situation is more similar to real cases and so allows to understand if the methods can be effectively adopted in the considered applications.

6.5.2 Basler blaze 101

For the fact that Basler blaze doesn't provide auto exposure procedures, it's necessary to adapt manually exposure time. In this camera, exposure time is directly linked to both intensity image and depth map quality. By lowering the exposure time to increase intensity image quality and avoid motion blur effects, the depth map gets worse and worse. This phenomenon happens since depth measures are based on light captured by the sensor. If a low quantity of light comes back to the image then the depth measures will be approximative and oscillating.

Since ND filters have been used to mitigate the too high laser power for this low distance application, when the exposure is decreased the filters can be lowered in darkness to have more or less the same degree of lighting in the images.

Another problem arising with low exposure times deals with depth estimation quality. Lowering the exposure times it's more difficult by the camera to reconstruct precise depth of points and oscillation of depth esti-

mates increases.

Basler intensity images are more blurred and bad illuminated due to two principal reasons. The first one is the presence of an additional, not intended, Neutral Density filter that compromises both focus and depth estimation of the camera influencing rays intensities coming back to the camera. In general, an additional lens in front of the camera lenses compromises the focus quality and, if it's dark, lowers the amount of light collected by the camera in a certain time. By calculating depth through the Equation 2.1 in Paragraph 2.1, the expected light intensity coming back from a certain distance has to be lowered respect to the one considered in the equation. For this reason, the depth map calculated from the camera is no more reliable. A solution could be the one of calculating the mapping error due to this filter based on the conveyor ground truth and the measure provided by the camera and rescaling all the depth measures by this error.

ND filters may be eliminated with objects that are not reflective and lucid, with these objects the laser lightning doesn't affect so much the depth mapping and local saturations are avoided. Clearly, since the images are collected near the conveyor belt (about 30 cm), the exposure time without ND filters have to be kept at its minimum (100 μ s).

The second cause of the bad intensity image quality deals with low exposure times set in order to avoid local saturations in objects. Local saturations in objects are strictly dependent on their materials and cause incomplete depth object mapping and, as a consequence, a not good performance of the algorithms designed. The solution adopted based on filling objects with post acquisition elaboration based on median neighbor values are efficient but don't give precise and real results. They are, of course, better with respect to having holes in images but the depth map obtained is far from the ground truth one.

MLP Method

The phenomenon described in the last section of the previous paragraph dealing with oscillation of global depth estimates is very negative for the MLP Method applications since a classification network calibrated for one depth image could be not valid for other images. Global pixels values oscillations compromise the classification procedure and makes this method unusable with moving images. MLP is not flexible to adapt to global oscillation, instead it's very strong in compensating local oscillations of depth values. In static cases, in fact, the exposure time was kept higher and there were only local oscillations. An example of global oscillation phenomenon and the consequent bad mapping is shown in Figures below.

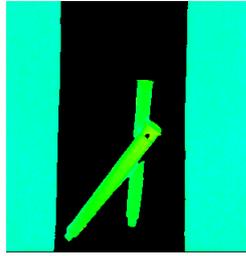


Figure 6.74: Classification of Training Image

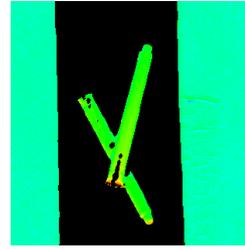


Figure 6.75: Classification of Depth Image

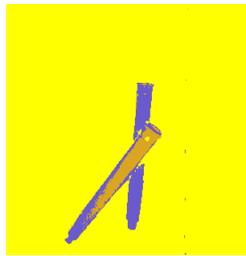


Figure 6.76: Training Depth Image



Figure 6.77: Test Depth Image

The upper left image has been used for the classification and the correspondent good classification is shown in the bottom left one.

The top right depth image is color different from left right even if objects are in the same overlapping relation. The classification colors in bottom right image should be others respect to the ones obtained and should be more similar to the bottom left one. Occluded object, for example, should contain less goldenrod color and more violet one.

The explanation of this results is that depth colors in top right image are slightly shifted in intensity and this causes the wrong classification. Thresholds estimated by the MLP training, in fact, are static and aren't able to adapt to this kind of global depth shifts.

A filter that reduces this effect is of course the Temporal Filter. By activating this filter, depth estimates get more stable and compatible with the MLP method. However a filter of this type is destructive in terms of depth estimation in moving objects. The idea behind this filter, in fact, is to collect and base depth estimates also on previous estimates. Since objects move, using this filter gives estimation partially or completely wrong. An approach of this kind would estimate current depth values with depth values coming from situations in which objects were in another place in the image. The idea of using this filter has to be discarded and the usage for alternative algorithms is necessary.

Derivative and Common Area methods are less sensible to depth local oscillations and could provide good results.

Derivative Method

The main problem arising with moving objects deals with bad depth estimation of final parts of the objects. An example of this phenomenon is shown in Figure 6.78.

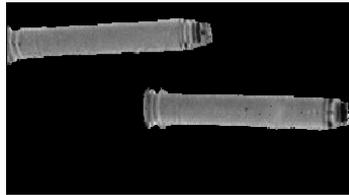


Figure 6.78: Free object Depth map (blue) and its derivative (red)

Only the right one of these two objects is examined since the other one is crossing the image border so it is discarded. Due to movement, depth is difficult to be mapped especially in the object's final part and this causes to have bad and not reliable estimates in objects' terminal parts. The derivative plot of the object is shown in Figure 6.79. A peak of the derivative, reflecting depth mapping discontinuities, is present on the left part of the derivative plot. This object should not present high depth variations since it's a free object.

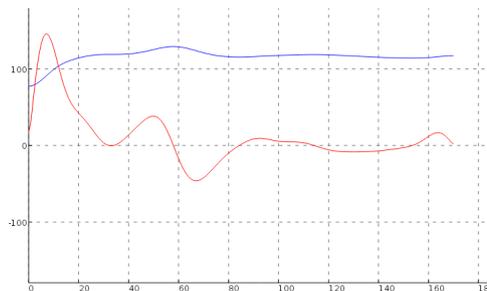


Figure 6.79: Free object Depth map (blue) and its derivative (red)

All the objects presenting this problem are discarded since their derivatives peaks are over the threshold imposed. For this reason, the recall parameter diminishes significantly and a lot of free isolated objects could be discarded only because of depth mapping problems. Looking at Figure 6.53 in which was presented a free standstill object depth trend, it's possible to see that the phenomenon just described didn't happen. Since this problem

is very difficult to be solved and causes the recall value to fall down, the Derivative method cannot be used for moving objects.

Common Area

The only stable method that could be used with Basler camera, in case of moving objects, is the Common Area one.

First of all, it must be remembered that the conditions to adopt this method are very strict. All the objects present in the intensity image have to be found by the search algorithm and depth holes in images have to be filled. With interactions of at most two objects, like in these cases, the identification algorithm is able to easily detect the objects. A big help also comes from the search area filtering that allows to keep lower the matching threshold without falling into shape matching false positives.

The search of the model with this particular object that is easy to orient having its contour shape, can be done both in the depth filtered image (when background is eliminated by confidence thresholding) and in the intensity image. Confidence thresholding consists in filtering away the parts of the image that are estimated with low confidence, in this case the background that has a dark color can be easily eliminated. The research, in the filtered depth image, gives faster results with respect to intensity search and is more suitable to have a precise object orientation without the effect of motion blur (that could be present in the intensity image).

Respect of the recall values typical of the analysis with depth holes, with the filling holes procedure designed the recall value increases evidently and the errors caused by holes are eliminated.

An example of the ordering result obtained with white untextured syringe is shown in Figure 6.80. The green contours means objects have been identified in that position by the shape based matching algorithm. Tests confirming results with this product have been done on 21 samples among a total of 83.



Figure 6.80: Common Area Method output order

Black textured syringe

As told several times in this work the Basler camera fatigue to map in depth black objects. In the case of the syringe considered in this paragraph (Figure 6.81), only small dark textures are present in the horizontal object. It's correspondent depth image is represented in Figure 6.82. This images have been collected in static conditions without temporal filter activated. Small dark texture doesn't show to influence the integrity of the depth map in any way. From now on the focus will be on the horizontal syringe, the inclined one will be excluded from next considerations and analysis.

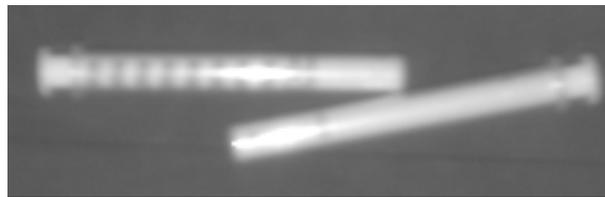


Figure 6.81: Intensity Image

Considering the same object (the horizontal one) moving on the conveyor belt (Figure 6.83), the depth map results very different from the one shown before. In particular, in correspondence of the black texture the depth estimation is wrong (blue instead of green). Depth estimated is more far respect the true one confirming the fact that dark parts may be difficult to be mapped. The dark parts absorb more light causing light reflected to be few, for this reason blue stripes arises where black texture is. As a consequence the obtained depth estimate is more far.

In Figures 6.84 and 6.85 are respectively shown the two 2D depth and derivative plots. While the depth map of the standstill object is almost flat, with low derivatives magnitudes, the moving object depth map and its relative derivative evidently oscillates. The results obtained in case of Derivative method is very bad since high oscillations would cause the major part of free moving objects to be discarded.

The oriented object doesn't show the depth distortion mapping just described since, as can be seen from Figure 6.81, it doesn't contain texture in the side shown to the camera. This is a double check of the fact that dark texture is the cause of this bad mappings.

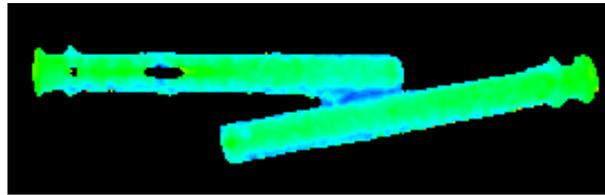


Figure 6.82: Standstill Objects Depth Image

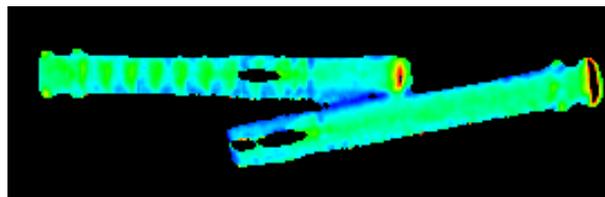


Figure 6.83: Moving Objects Depth Image

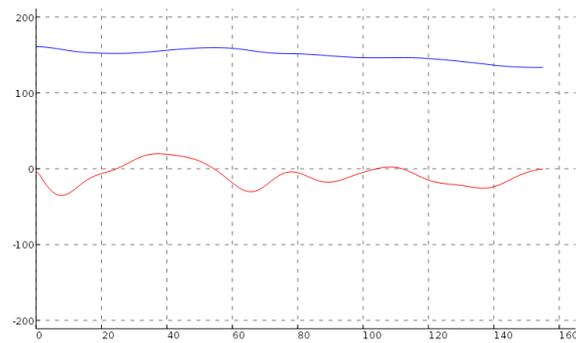


Figure 6.84: Standstill object Depth map (blue) and its derivative (red)



Figure 6.85: Moving object Depth map (blue) and its derivative (red)

As a result the Derivative method also with these objects cannot be used. In addition, a depth map of moving objects, presenting the mapping problem just described, cannot be adopted even in case of using the Common Area method. The mean values calculated don't give reliable results.

The effect of bad depth mapping in the extremes of moving objects can be noticed also from Figure 6.83 in the right extremes of objects. Inside objects are present big black not mapped holes surrounded by yellow and red depth estimates. These estimates are far from truth one shown in Figure 6.82 and put in evidence the fact that Derivative method would discard these free objects since high depth discontinuities are present in both moving objects depth maps.

Long Syringe

Tests of the Common Area method with this kind of product have been done on 21 samples among a total of 68 images. Of course some of them didn't contain products or the contained products were crossing image edges and so they were discarded.

Long syringe has a very particular shape that makes difficult to establish its verse direction. In Figure 6.86 are represented the shape model contours and it's possible to notice that the verse of the orientation is determined only by a final oval contour part that is the contour of the plunger. For this reason, the research needs more time in order to detect the right verse of the model when a match is found.

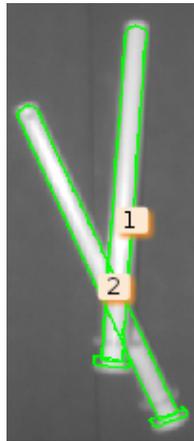


Figure 6.86: Common Area Method output order

Anyway among all images the 100% precision and recall rate have been reached and the method can be used also with this type of product.

This syringe doesn't contain black textures and no deterioration of the depth image quality is noticed. Remember that black textures may ruin depth maps when objects are moving, as shown in previous Paragraph.

The major advantage that the Common area method gives with Basler camera is that, in case of a good design of the search algorithm, complex objects overlappings can be managed. This algorithm is able to solve the classic "Shanghai problem" in which the global pick up order of objects is fundamental in order to pick up all objects without perturbing other ones. Some results are shown in the following Figures to demonstrate the working of the method.

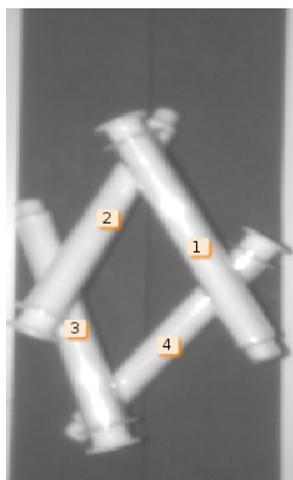


Figure 6.87: Global picking order

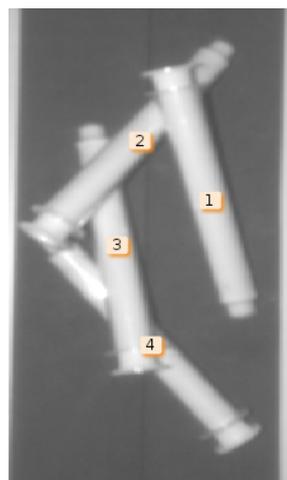


Figure 6.88: Global picking order

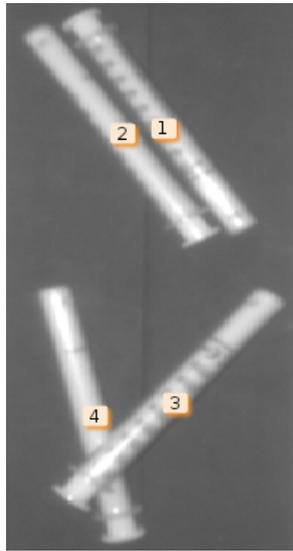


Figure 6.89: Global picking order

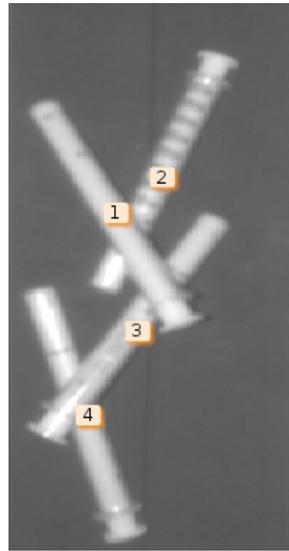


Figure 6.90: Global picking order

Timings

In case of Dynamics tests with Basler camera, some qualitative results about white untextured and long syringes in terms of timings are presented in Table 6.9. Times measured are not the effective time needed by the machine for the identification process. Numbers shown are related to measures affected by processes preemptions and memory accesses and so considerations can be done only in a qualitative and rounded up way. More details about timings measurements are reported in Paragraph 6.2 in the "Computational timing" section.

	Search Time	Computation time	Overall Time
White syringe	36,3	25	61,3
Long syringe	31,9	23	54,9

Table 6.9: Common Area Method Timings in ms

In the Computation column, timings for the processing of two objects are presented. If more objects are present in the image then the timing arises significantly since all combinations of shapes interceptions have to be found to order objects. As said before these timings vary case by case and so only estimates to give the time dimension order of processes can be done.

At the Total times shown, the exposure time has to be added jointly with the Filling procedure time that needs few milliseconds to be done.

The main consideration that can be done is the same as the Intel RealSense case: the search time of all objects covers a very significant part of the overall computation time. In the next Paragraph some consideration about shape matching computation time reduction will be done.

6.6 Some specific cases

Among the objects analyzed some of them aren't compatible with identification methods described until now. It's the case of small pipettes or small tubes that have the problem of not having the same thickness along their axis. With objects which have not the same height along their axis or are thin, the analysis of overlapping cases described in Paragraphs 5.4, 5.5, 5.6 cannot be adopted, mainly for the fact that these methods detect overlaps based on depth variations. If an object contains these variations intrinsically in its depth map then these methods become useless. Also having objects height dimension similar to the camera estimation error make it difficult to distinguish objects with respect to foreground or changes in height due to overlappings. As is possible to observe from Figure 8.67 it's difficult to detect height changes of one tube caps over the other. In this case height doesn't allow to establish which object is upper respect to another and, apparently, the depth image is useless in this detection problem. The example comes from the Intel camera but also the Basler fatigue to map these low difference in depths. The objects that are mapped in depth are the ones represented in Figure 6.89 .



Figure 6.91: Pipettes Overlapping Depth Image

Derivative method and MLP method cannot be adopted since depth variations in an object's profile wouldn't make them able to detect difference in height caused by an overlapping or caused by the object profile itself. Since objects are thick, especially in their caps, unreliable depth maps may be generated due to partially wrong estimates in the thickest parts. Remember that, especially around contours, the Realsense camera has uncertainties in estimating depth.

The Common area method relies on area interceptions that, in the kind of objects considered, consists in very small areas. The analysis of them are then not so reliable due to their dimension.

In conclusion, these three methods are weak in detecting free objects when their shape is not thick enough or their depth varies along their profiles.

Anyway some considerations to detect free objects can be done with images collected. Using the output coming from the shape matching algorithm all the positions of objects are found in the image. Checking the interceptions between each couple of shapes found, it's possible to detect overlaps:

- If the interception area differs from zero both objects are discarded because no algorithm can be used to detect which is free. In general the area of these interceptions must be significantly different from zero. Interceptions made of few pixels may be caused by near objects that touch themselves. In those cases both objects can be collected and have not to be discarded.
- Being in the situation in which the interception area is significantly different from zero and it's not possible to detect which object is free and which occluded doesn't preclude detecting free isolated objects. If the area of contact is consistent (so major of an established value) then the only thing that could be done is to discard both objects since there's no possibility to pick up one of them being sure it's free. In case their interception is null objects are examined and collected.

It's important to underline that depth image is not completely useful in these cases since it carries information about where the objects are. By thresholding depth image through a manual thresholding or a Multi-Layer Perceptrons network classification is possible to establish which part of the image belongs to foreground and which at the background. As introduced in Paragraph 5.3.1, a binary mask can be derived from depth information.

In Intel RealSense camera, the approximative area in which objects lays is extracted by one of the three depth channels. By using this binary mask it's possible, then, to filter out a region in the RGB image in which search for the objects. It's always necessary to dilate the foreground mask in order to generate a Search Image for the shape matching since depth is mapped wrongly in Intel RealSense cameras.

The Basler camera provides better precision and correspondence between mapped contours and real ones. This allows to extract from the depth image directly the precise area in which objects are. Anyway the shape matching algorithm needs also parts around the contours in order to get gradient informations and detect contours so, even in case of the Basler camera, it's necessary a smaller dilation with respect to the Intel RealSense one. Dilation

parameters of the masks depend strictly on the objects chosen, they have to be selected and adapted case by case.

This procedure of filtering the search area reduces remarkably the search time and avoids having false positives along search. These results are evident in Figure 6.92 where are shown timings of shape searching for filtered and not filtered RGB images of a pipette. In mean the computation timing to search an image between a 640x480 RGB image and its filtered counterpart is reduced by 43.7%. Images used contained two objects each one.

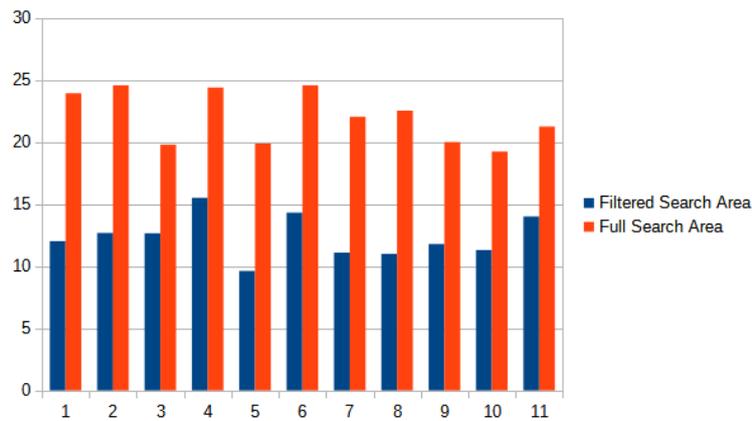


Figure 6.92: Full Search Research (red) vs Filtered Research (blue) Timings

As the technologies used give different advantages and disadvantages, the method used for the shape matching may vary. In particular in the Intel Realsense camera the search can be done only in the RGB image since the depth image is not precise and contours map varies object by object. The shape matching algorithm is not flexible enough to search in case of these variations. Additionally the depth and RGB images do not precisely correspond and finding an object in the depth image would not give the right position of the object in the RGB image and so in the space. Anyway RGB images are well detailed and allow to get good matching results even with perspective distortion cases. In Figure 6.93 an example of matching result is shown. Shapes matched are shown with green contours, the degree of precision in matching and orientation is high.

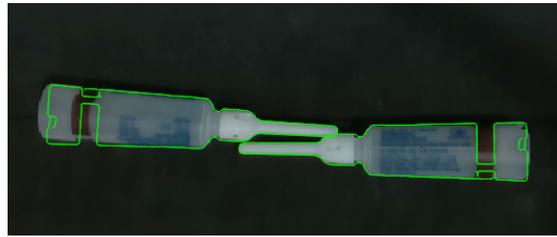


Figure 6.93: Pipettes RGB Image with shapes matched highlighted (green)

Basler provides, instead, a depth image that is aligned with the intensity one. An eventual identification of the products in this image would lead to the same results obtained with an identification in the intensity image. Anyway, searches of objects model in the depth image is not very efficient since objects are characterized by different colors depending on their position and height so a model carried out with a certain object depth model would be found with difficulty in generic depth images. In order to be aware of local and global contrast variations in objects the "ignore global polarity" or "ignore local polarity" search setting should be used. These two settings would make greater the research but with the drawback of increasing remarkably the computation times.

A better solution coming from the usage of both lucid and opaque black conveyor belts is the possibility to detect and exclude background from foreground with a simple threshold of the confidence plot. A binary threshold can be applied giving the possibility to have a binary mask containing only part of the image that belongs to the foreground. Around the conveyor belt there could be other parts contained in the binary mask but these will be easily filtered away reducing the image domain only at parts in which the conveyor passes. An example of this mask is represented in Figure 6.94.

Now, creating a model for the object in the binary image, the research can be done in these binary masks with respect of doing them in the intensity images. The advantage of searching an object in a binary image is that the search is very very fast, lightweight and precise. No disturbance objects come from the environment that may compromise the mask creation or the objects depth mapping. The detection using binary masks gives an additional advantage connected to precision of the shape orientation and positioning. In particular these masks contain only points belonging to objects and so the orientation comes directly from interpolating the model with the mask blobs.

The threshold for the research can be kept lower respect to search in the intensity images since if an object is matched the possibility of being a wrong match are very low. In other words, in the mask white parts belong of course to an object that must be found. Instead, in dark parts, objects won't be matched.

Clearly, this kind of identification doesn't take into account the fact of having, for example, perspective projection problems. In particular objects deformed by perspective projection may create very deformed binary shapes that may confuse the search algorithm. In case of perspective projection problems with binary masks the intensity image can be always used to search for the objects.

In general, the research with perspective projection deformation gives better results in intensity images with respect to binary masks. In fact, the research for deformed objects may make use for the identification also on details and shadows inside the objects. Binary masks are white and plain giving no information on the internal curvatures or details of objects. Intensity research result is shown in Figure 6.94 . With pipettes and small tubes the differences in search results between intensity images and binary masks are the same. The only difference is that certain objects distorted by perspective the identification in binary masks may give bad results. On the other hand the binary research for the major part of object gives more precision and velocity in the research task. In conclusion the binary research is a valid alternative to search for objects. It has been used also with the white untextured syringe in previous Paragraph (6.5.2). For other syringes types the binary mask was not enough well defined to grant orientation of shapes when matched.



Figure 6.94: Binary mask of objects in Image 6.93

Another fact to consider, that could allow to be independent from a commercial library for shape matching algorithm, is that the binary mask can be directly used to orient objects. By filtering out the blobs that have areas bigger than the area of one objects the isolated objects blobs are obtained. Thanks to moments the barycentre, orientation and verse of these blobs can be established having in a clean and fast way the information needed by the robot arm to process products. No evaluation on touching objects that can be grabbed but form a big blob could be done no more.

In case of the example of Figure 6.94, two of that three objects could be picked safely, the third one have to be discarded since may be moved from the picking up procedure of the near one. By not using Halcon shape matching procedures the two right objects cannot be detected since their

blob is discarded and only the right one is collected. Using this method may decrease the efficiency of the algorithm but can be a valid alternative to be independent from a commercial library. An example of the application of this procedure is shown in Figures 6.95 and 5.6 . The usage of a V section shape conveyor belt may help a lot in dividing and isolating objects in order to increase efficiency.

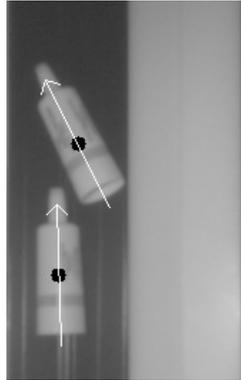


Figure 6.95: Object orientations result through binary mask and moments

An important problem with pipettes and tubes analyzed is that there is the possibility to have one of them embedded into another one. Since the major part of them are open in one side, this phenomenon is very frequent. It's not granted in these cases that both of the pipettes are grabbed by the robot arm since they could be not separable or the robot arm could not be able to grab one pipette since another is partially inside of it and the vacuum cap force is not strong enough to catch it. Both of these cases have to be managed, the solution adopted is to discard cases in which this phenomenon happens and so when the interception of the two objects blob is evidently different from zero. The embedding situation is managed as the overlapping one, so no additional routines have to be added. An example of embedding situation in which objects are discarded is shown in Figure 6.96.

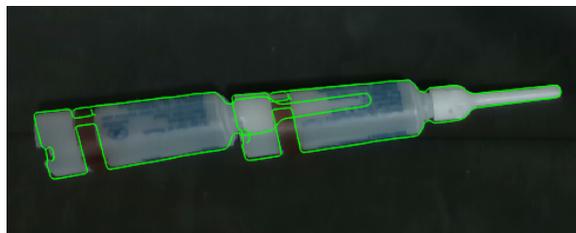


Figure 6.96: Example of discarded embedded objects with matched shapes (green)

In case two objects are touching but they are not occluding each other,

only one of them can be collected since the grab procedure of it may move the other ones compromising their positions. In these situations only one object is collected in binary blobs containing more than one object. Blobs containing more than one object are found knowing more or less the pixel area of a free object. If these objects are all free but only one of them is grabbed the other one contributes to decrease the performance of the algorithm. The problem just described may be solved using V section conveyor belts that separates automatically objects and avoid having them grouped together.

In Figures 6.97, 6.98 and 6.99 three examples of the problem just described are shown. On the objects to be collected is written the word pick otherwise they are discarded. In Figure 6.99 the shape matching result is highlighted by green contours. Also in these cases, high precision in matching is reached.

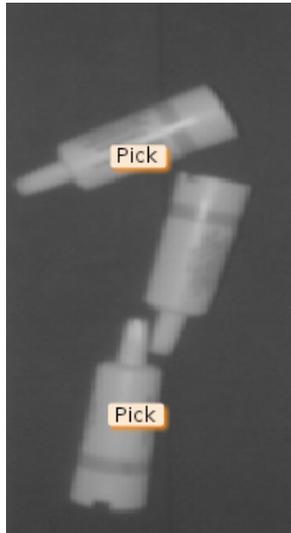


Figure 6.97: Touching objects management case

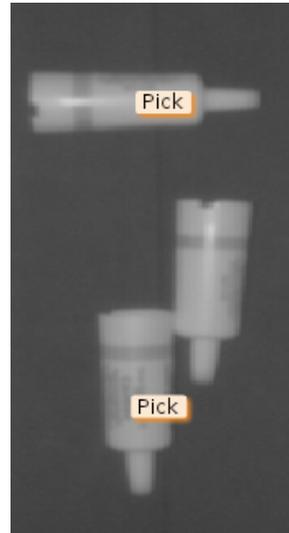


Figure 6.98: Touching objects management case

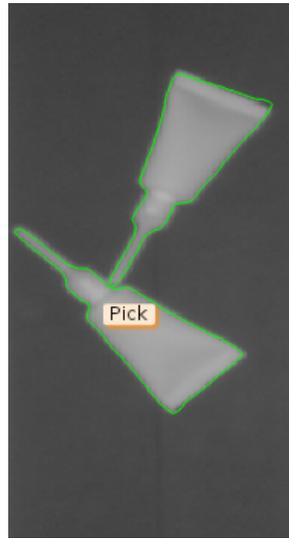


Figure 6.99: Touching objects management case

Major part of the objects now considered are very prone to roll away during the movement of the conveyor belt. The adoption of a V section conveyor belt is so essential to avoid products rolling when they are moving on the conveyor.

Respect of what anticipated in the introduction of this section, the choice of a V section conveyor belt may generate problems for some objects. In particular looking at Figure 6.100 the barycentre and orientation of the objects it's compromised by the inclination of objects caused by the shape of the V shape drawers. In this particular case, the V shape drawer has to be avoided since these objects may lie inclined on the conveyor causing to have not accepted barycentre and orientations of them. This phenomenon of bad positioning may be observed by comparing the different objects shapes that are not symmetrical and equal each other.

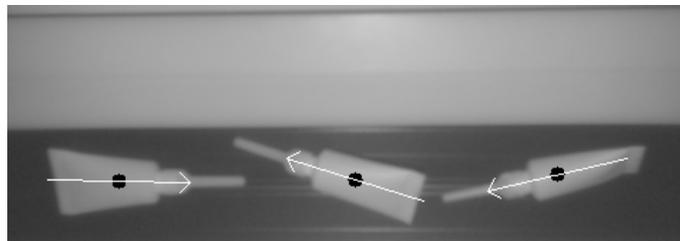


Figure 6.100: Inclined objects orientation case

The solution is to use, only in this case, a plane dark conveyor belt in order to orient objects parallel to the camera all in the same way. These

objects are not cylindrical but have a plane shape that prevents them from rolling on the conveyor belt. Some results are shown in Figures 6.99 and 6.101.

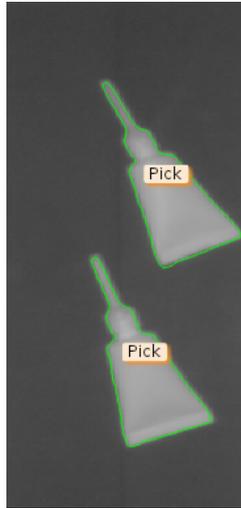


Figure 6.101: Plain conveyor belt case

This object type demonstrates another time that the conveyor belt choice is fundamental and essential for the best working of an identification algorithm and that the best one has to be found for each specific product depending on its shape and color.

Timings

Timings of described techniques will be provided ahead in examples of detection of specific objects. As told several times, the data presented are only indicative, to give an idea of computation timings of described methods. Algorithms based only on single objects blob detection in binary masks require less time than the one required for these examples since no shape matching is needed. The timings are very low respect to the ones shown in Tables 6.7 , 6.8 and 6.9 due to fact the shape matching time is evidently reduced. Depending on best results given, the intensity or binary confidence images have been used for the shape based matching. The timings aren't very different for searches in intensity or binary confidence images. Static and Dynamic words between parenthesis mean that tests have been done on standstill or moving conveyor belts. Timings are related to images containing two objects to be identified.

	Search Time	Computation time	Overall Time
Pipette in Fig. 6.99 (Static)	1,3	2,5	3,8
Pipette in Fig. 6.96 (Static)	1,0	8	9
Pipette in Fig. 6.98 (Dynamic)	1,0	11	12

Table 6.10: Timings of algorithms applied at binary images in ms

Dynamic tests

Basler camera gave problems in mapping in depth objects moving on the conveyor belt and, since the analysis on binary masks is very efficient and appears to give interesting results, some tests have been done in case of moving pipettes. The type considered is the one of Figure 6.92 . As always done in dynamic cases a filter for objects crossing the image have been implemented in order to consider only fully included objects in the image.

Tests on images demonstrated that the method based on binary mask detection grants still precise and good results even in dynamic cases. Tests also stressed cases of near objects in which only one of them have to be kept and cases of partially embedded objects that are discarded.

Chapter 7

Conclusions

In this work, the cameras used to accomplish the identification tasks are the Intel RealSense D435 and the Basler blaze 101. They are very different cameras since they rely on two different depth mapping technologies. The Intel one is an advanced stereo camera that is able to provide depth images through the matching of left and right images with the additional help of a randomly generated infrared light pattern projected in the field.

The Basler camera relies, instead, on the depth calculation through time of flight principle: depth is calculated by evaluating the time needed by a light beam projected to come back to the camera.

These 3D cameras could grant interesting improvements in terms of productivity and performances especially with certain kinds of objects. Satisfactory depth images coupled with point cloud maps can be obtained from both the cameras. Depth and color/mono images can be successfully used to make evaluations about overlapping objects and pick up order. These considerations would result difficult or even impossible by using 2D cameras considering also the fact that even establishing which object is higher respect to the other, its height wouldn't be known.

7.1 Images Quality

Thanks to the usage of black conveyor belts both the cameras can extract useful depth and RGB/intensity images. It grants high contrast and low evidence of presence of shadows caused by ambient illumination.

Both cameras provide filters and settings to be adaptable to each different situation, Intel RealSense in particular provides higher degrees of freedom like in selecting color maps criterion, exposure time ranges, saturation, sharpness, IR pattern intensity and gain parameters.

7.1.1 Intel RealSense D435

Intel Realsense provides a sharp and detailed RGB image even in cases of moving objects with Gain increases and major image noise. The filtered RGB image is better in terms of quality both in Static and Dynamic cases with respect to Basler: the shape matching procedure works better with color images and Intel camera images are sharper and with more contrast. The overall matching procedure results faster and more precise respect to the ToF camera.

Depth images are collected using High Density preset giving low confidence values in the matching procedure of the stereo images patterns. The Filling Rate of the image is higher but depth estimates could be less reliable especially around contours. In order to increase the Filling Rate the "DS Second Peak Threshold" have been further lowered. Additionally, to have 100% Rate the Holes filling post processing filter has been activated. It keeps about 3 milliseconds of computation time but gives a dense described depth image. All algorithms are very sensitive to depth holes in images and so, if possible, they have to be eliminated.

The depth module parameters don't need to be reset for moving objects situations, only small and not invasive changes in parameters have to be done for the RGB module in order to avoid motion blur.

7.1.2 Basler blaze 101

The intensity image quality is worse with respect to the one provided by Intel camera. It's in grayscale instead of RGB format and it's more blurred. The blurred and not sharp effect of provided images comes from the Neutral Density filters adopted to mitigate the high intensity of the infrared laser that illuminates the scene for the camera. This camera, in fact, makes use of it's proper source of light and it's immune to environment source of light thanks to its pass band filter. ND filter allows to collect images without oversaturation effects in them. One of the problems of this camera is, in fact, the mapping of high reflective objects that may present local light saturations and consequent holes in the depth image. A trade off between exposure time and darkness of ND filters have to be selected in order to avoid motion blur and have well illuminated scenes.

The ND filter, since not foregone by the camera designers, compromises the depth estimates reducing light intensity coming back to the camera sensor. The consequent depth calculation is no more reliable.

In terms of parameters the camera gives high customization freedom, both depth mapping range and exposure time can be set independently and in a high range. The main and only drawback is that the light source intensity cannot be reduced in intensity and it's necessary to use ND filters to mitigate its effect.

The main problem given by the ToF cameras and that is difficult to counteract is the multiple reflection problem. Since objects considered in these analyses are quite thick, a consistent part of them may be badly estimated. Dark conveyor belts help counteracting this bad behaviour absorbing multiple reflected light beams.

Another problem given by this camera is that lowering the exposure time the camera gives more oscillating depth estimates and compromises the usage of algorithms. This problem arises especially in dynamic cases when low exposures times are needed to counteract motion blur effects in the intensity image.

A fast and easy segmentation can be provided directly from the confidence image thresholded during the image acquisition. By exploiting the fact of having black opaque conveyor belt behind the object, the camera maps the background with low confidence. In this way, using a threshold value around 6000, that is a low value considering it ranges between 0 and 65535, the confidence map provides objects blobs segmented for free. This process can be done only with dark backgrounds.

7.2 Algorithms Designed

Connecting contours and Background subtraction methods shown to be weak in terms of precision and completeness of estimates. Contours detection is prone to errors in the precise reconstruction of precise objects shapes. Contours connection, in fact, may badly merge the contours leading to consequent wrong barycenter and orientation estimates.

Background subtraction is a technique adoptable only with Basler camera since the depth map of Intel camera is very imprecise. It allows to be aware of the height of the conveyor belt in which objects lays and it's resilient also to situations in which the camera is not perfectly parallel to the conveyor belt. Hence analysis can be done in an absolute depth framework in which only the magnitude variations between background and foreground counts.

Anyway this procedure gives the possibility only to detect isolated objects and not to analyze possible overlapping situations.

The three final algorithms proposed demonstrate to be flexible enough to be easily adapted with various objects and to cover possible eventual camera problems. A brief resume of them can be done:

1. Derivative Analysis

This method is based on the extraction and analysis of depth pixels values inside each object major axis. In particular, when an object is detected, depth values along its axis are examined in terms of their derivative peaks. Derivative peaks are indexes of overlapping objects since high depth variations means something is present on the current object. Thanks to a threshold on the peaks values the object is classified between grabbable and not grabbable. Their orientation is mainly based on the calculation of the standard deviation inside axis values. If it's low they are horizontal, if it's high they are oriented.

The main advantage of this method is that it works even if not all the objects present in the scene are detected. Grasping decision is based on the depth information of the detected objects. Other object detection or positioning doesn't influence the picking decision. This is a big advantage since if objects are difficult to be detected and no shape matching false positives have to be found the threshold of the search algorithm can be set high without worries about the fact that research couldn't find all of them.

This method relies on the information inside all the object axis that could be compromised by bad local depth estimates or depth holes. This could increase the recall value decreasing the efficiency of the algorithm. Remember that, in this work, efficiency is intended as the ratio between grabbed and grabbable objects.

2. Multi-Layer Perceptron

This method is based on the analysis and classification of the depth image values dividing them in three categories: High, Low and Background. The classification is done automatically through a Multi-Layer Perceptron neural network opportunely trained before the classification. This speeds up the selection of depth ranges in which classify depths and gives more robust and flexible results. In addition the training procedure is very easy and needs only a sample area for each class. MLP is the best network in order to have fastest classification speed, offline training phases and low memory requirements.

The Background areas are used to know where, for certain, objects have not to be searched. The Low and the High classified areas are used instead to detect free and/or occluded objects. In synthesis free objects are detected since they don't include High classified parts. If, instead, they contain large parts of High areas they are classified as free and oriented. In other cases they are classified as occluded.

This method doesn't require all objects in the scene to be found if and only if the overlapping situations includes at most two objects. In that case interactions of more than two objects are not considered. This is the case of the major part of objects combination in the application considered so not a large decrease of the recall value will be experienced.

Otherwise also the cases of interaction of three objects may be included but with the hard requirement of identifying all objects in the scene. Depending on the usage and on the context of this algorithm one or the other method can be used. The orientation can be estimated based on the depth variance inside objects or on the included High area inside the found shapes.

3. Common Area

The Common Area Method relies on the analysis of common areas shared by the overlapping objects. Areas are determined considering the objects two by two. Thanks to the analysis of the mean values of depth pixels stripes oriented in the same orientation of the objects considered makes it possible to establish a height hierarchy. Through this hierarchy is possible to establish which objects are free and which are occluded. For each object it can be also told it's pick up ordering based on how much objects are occluding it.

Main drawback of this method is that it requires all objects inside the examined image to be found in order to work properly. Otherwise the pick up ordering and the free top objects cannot be determined. The search area reduction gives the possibility both to reduce the arising of shape matching false positives and to keep low the threshold for the search algorithm. In this way all objects can be detected easily.

In case of holes presence inside the depth images also this method suffers since holes may arise inside the common area leading to wrong results.

This algorithm, on the other hand, gives more information about objects configuration since allows to establish not only the free objects but also the overall pick up order of the occluded ones. Anyway, in this context, this results isn't very useful since it's enough to know the top free objects positioning.

7.2.1 Basler blaze 101

Considering the Basler camera it's possible to say that actually it's best usage is at medium far distance from targets. For near usage, in fact, it's

required to mitigate the high laser intensity with additional dark filters that makes the camera still working but with lower image quality performances. A good application, for example, could be the one of palletizer machines. In that case objects are big and the problems of bad contours depth estimation coming from multiple laser beam reflection is negligible respect to objects dimensions.

In case of our application, the small dimensions of the objects ruin more depth image especially near object contours. This effect can be mitigated by the usage of dark opaque conveyor belts that carry advantages both in multiple reflection avoidance and in helping reduce the search area. As explained in the previous Chapter, thanks to the confidence image, the search area reduction comes almost for free. Objects that are transparent cannot be detected by this camera, lucid ones instead may present problems of local holes in depth maps due to light saturations.

Corrupted images may be correctly recovered thanks to median filters that don't impact too much computation times (few milliseconds) and allow all algorithms to be used.

With static images this camera gives very good results, it grants higher depth precision result respect to Intel Realsense representing also objects curvatures. In particular thanks to the usage of Temporal filter the quality and stability of images increases evidently. It's possible, after the filling holes procedure, to apply all the three designed algorithms.

Since depth estimates don't oscillate no more, the classification network can be trained in a stable way and the Multi-Layer Perceptron method can be applied without encountering problems. Clearly, the drawback of the number of objects to be managed remains of at most three objects.

In general the algorithm works safer with at most two objects interacting in an overlapping relation.

Also depth and correspondent derivative plots, provided by the application of the Derivative method, are very clear. The smoothing procedures are enough to obtain clear and evident peaks only when an overlapping occurs. Of course since the holes filling procedure is not able to perfectly reconstruct objects' depth map some false negatives could arise but in low quantity.

In the case of Common Area method it works properly even if closing of holes isn't done in a perfect way. It's the most insensible method to the filling hole quality among the proposed ones. One drawback of this method is that it needs all the objects to be found in the scene. If this strong constraint is satisfied then the recall and the precision values are for certain both at 100%. Especially when objects overlapping are low in number this constraint is respected.

Worse performances come from dynamic cases tests since depth estimation errors arise and the only algorithm that continues to work is the

Common area one. Objects contours parts become difficult to be depth mapped and errors in estimation arises (Figure 6.78, 6.79).

Common Area method still continues to work in cases in which few objects are contained. It grants 100 % recall and precision rate satisfying in the best way the requests imposed by problem constraint. It's of course still possible to know also the pick up ordering of objects. The overall detection procedure of free objects needs less than 100 ms. Specific timings for objects are shown in Table 6.9 .

With objects containing back textures, the depth map results highly compromised since black causes wrong and less reliable estimates. In these cases it's impossible to apply none of the three methods, even the Common Area doesn't work any more. In Paragraph 6.5.2 an example of black textured syringes is reported with it's related depth map.

Anyway 100 % recall rate is granted with different typology of objects permitting this camera to be competitive with other ones even in dynamic cases. Variable or low lightning doesn't constitute a problem since the camera uses its own laser source and is provided by a selective band pass filter in the invisible spectrum that filters out visible spectrum components. The exposure time has to be opportunely adapted to avoid motion blur.

7.2.2 Intel RealSense D435

This camera performs very well for the application kept in exam. After the depth-color images alignment procedure, one between Derivative method and Multi Layer Perceptron method can be used. The Common Area algorithm is not adoptable since the depth map is uncertain especially near objects contours.

Anyway both in static and dynamic cases the results are very satisfying. In static cases both MLP methods can be used with any problem. The methods have been tested also with very thin syringes or with syringes with not very linear sections (Big head syringes). In all of these extreme cases the algorithms demonstrated to continue to work properly.

In dynamic cases this camera is able to counteract better motion blur and to give higher quality depth maps respect to Basler one. Its performances in case of moving objects actually outperforms the Basler ones. Transparent objects can be seen and mapped in depth, even if with low accuracy, covering most of the cases in which Basler fails. In order to get a high depth map filling rate High Density preset and holes filing post acquisition filter are necessary. The depth image results less detailed than the Basler one in terms of object shape definition since Realsense is able to detect smaller details but with overall less precision. For these reasons it's impossible to extract only from depth images sharp object shapes from the depth image to detect the exact object barycentre position. Identification timings vary between a few tens of milliseconds to more than 300 ms. The major part of

the algorithms timings is affected by search timings. Based on the search of the model complexity and of the degree of perspective deformations the computation time increases. The MLP classification is very light weight, it lasts about 10 milliseconds in each object. For all objects it's granted to have 100 % precision and recall values going from 95 % to 100 % . Isolated objects have always a recall value of 100 %. Tests have been done in large data sets to demonstrate better the working of these two methods. Respect to the Common Area method it's not mandatory in these cases to find all objects in the scenes. In Tables 6.1 to 6.8 are respectively shown timings and detection statistics.

The only problem to be solved remains the illumination one since a not uniform illumination can compromise camera performances. The usual procedure to be independent of it is to use auxiliary scene illuminators.

It's possible to conclude that this camera outperforms the Basler one in terms of price and performances since it's able to manage an higher quantity of heterogeneous objects in both static and moving conditions with very low false positive rates. The Basler camera costs about five times the Intel one.

Computation timings are low enough to make the robot able to know in time objects orientation and position for the grasping procedure.

Further code optimization could be done to reduce even more computation times and make algorithms compatible with real machines. Remember that computation times provided are highly qualitative but useful an idea of the magnitude of the computation times required for the algorithm's execution.

The search process requires the major part of the time spent in computations. It could be reduced significantly before the matching procedure when areas not including objects are filtered away using depth information or Multi-Layer Perceptron method. It has been demonstrated that searching in the whole area with respect to searching in the reduced one could reduce search times up to 51.5 %. The search area filtering is the first step that could be done as optimization. Clearly, search algorithms and the searched model creation could be improved even with real tests with robot arms in order to establish best parameters configuration to give both enough precision for the catching procedure and lightweight computations.

Finally designed algorithm code could be also improved by coding more efficiently their workflows. Code design was mainly devoted to check and demonstrate the concept and the working of algorithms.

The designed way to discard objects that crosses the image borders is able to filter only objects fully included in frames grabbed. Remember that this avoids problems of wrong shape matching and/or allows to exclude

objects from further analysis since it's not for certain known if they are occluded or not.

Except for the case of Long Syringes that need the cameras to be a bit more far from the distance at which tests have been done (about 30 cm), both cameras are granted to have at least one frame in which objects are not crossing the image borders and are so exploitable by algorithms. It can be concluded from this point of view that for objects tested the distance chosen is the right trade off between detecting objects at least in one frame collected and having the object big enough to search and match them in a satisfying way.

In case of Long Syringes the cameras should be put about 10/20 cm far from the conveyor belt in order to have best identification results.

In general one important part of the camera's setup, apart from its parallelism respect to the conveyor belt plane, is its distance from the conveyor belt. If it's too far objects would be too small to be finely identified, oriented and examined; if it's too near objects risk to be excluded from the analysis since they are always crossing image borders.

It can be concluded that at least one of these detection methods can be used to detect free objects being aware of having false positives with both Intel and Basler cameras. False negatives could arise but in low quantities giving very high recall rates and always 100% precision (as imposed by problem constraints). The 100% of free isolated objects are detected and almost the 100% of them are classified as graspable. The right camera has to be chosen depending on both the application and the characteristics of the object to be treated.

7.3 Final Considerations

If objects are small or thick both cameras allow anyway to filter out the search area for the search giving advantages in computation speed up and research quality respect to 2D cameras.

As explained in Paragraph 6.6, small overlapping areas of objects or objects that have no linear sections are difficult to be managed by advanced algorithms. In these cases, only near objects, not lying on others, can be classified as grabbable.

The only thing at which it's important to pay attention is that too near objects may partially touch other ones. In these cases the algorithms are able to select only one object among them to be picked up. Of course the algorithms' efficiency diminishes but no catching problems arise due to fact objects are moved from positions in which they were by the grabbing of objects in contact with them.

The presence of depth images have not to be underestimated since it's less color and texture independent. It may be a valid alternative for segmentation respect of using field images. As said before, depth images suffer from imprecisions around borders. Basler camera has sinking borders and Realsense camera is used with High Density preset to have a higher Filling rate. The drawback in Intel camera is that uncertain, not sharp and not precise contours are obtained.

With these objects the computation times are very fast since algorithms are based on the search of small objects in very reduced areas. In case of Basler camera, search timings are further reduced since the research could be implemented in binary images coming from binary thresholds of Confidence images. In this way few shape matching false positives arises and the precision in terms of positioning and orientation is increased. Tests with moving objects confirmed that these techniques are able to work both in Static and Dynamic conditions. In case of Basler camera, only few tenth of milliseconds are required to identify and classify free objects (Table 6.10) .

If the search algorithm won't be used at all, the binary image obtained from the Confidence one can be used with simple blob detection techniques. If a blob area is major of a single object approximate area then it's discarded otherwise it's analyzed. The orientation procedure cannot be provided by the shape matching algorithm since it's not used. Instead, the positioning, orientation and verse can be derived from blob moments calculations. In Equations 5.1 to 5.5 are shown the procedures to find them.

This procedure gives lower recall values respect to using a shape matching algorithm but can turn to be very useful when dealing with objects isolated, for example, thanks to a V section shape conveyor belt. It's a very lightweight identification procedure since it's based only on calculation of blobs parameters and it's also independent from the usage of a commercial library like Halcon. Only OpenCv library functions are enough to carry out results in a very efficient way.

Methods described in this paragraph can be applied to all the objects that cameras are able to map in depth. From the prefiltering of the images passing through the matching arriving at the identification process are procedures that can be always applied. Isolated objects can be easily detected and analyzed thanks to shape matching (Basler and Intel Camera) and to blob moments (only Basler camera).

From the identification point of view it's a very powerful result since these procedures are unique and standard for all objects kept in exam. For some of them also the three advanced algorithms described can be successfully used to detect free objects in overlapping relations. This is a very important result since the achievement requested (designing a method useful for the major part of objects considered) have been reached using advantages given only by 3D cameras.

Flowpack objects cannot be treated at all with these 3D cameras since depth maps result not reliable at all. Both of the technologies used are unable to map well and with high confidence transparent plastic parts due to local reflections of light.

Transparent objects, instead, can be well depth mapped but only by Intel camera. The Basler camera relies on laser projection to estimate depth and transparent objects are crossed by light without deflecting beams significantly.

Bibliography

- [1] Tri Khuong Sergey Dorodnicov Dave Tong Ofir Mulla Anders Grunnet-Jepsen, John Sweetser. Intel realsense self-calibration for d400 series depth cameras. https://dev.intelrealsense.com/docs/self-calibration-for-depth-cameras?_ga=2.241469509.632393114.1600067171-1040537290.1599567481.
- [2] John Woodfill Anders Grunnet-Jepsen, John N. Sweetser. Best-known-methods for tuning intel realsense d400 depth cameras for best performance. https://www.intel.com/content/dam/support/us/en/documents/emerging-technologies/intel-realsense-technology/BKMs_Tuning_RealSense_D4xx_Cam.pdf.
- [3] Intel RealSense D400 Series Product Family Datasheet. 01 2019. <https://www.intel.com/content/dam/support/us/en/documents/emerging-technologies/intel-realsense-technology/Intel-RealSense-D400-Series-Datasheet.pdf>.
- [4] D400 Series Visual Presets. 07 2019. <https://github.com/IntelRealSense/librealsense/wiki/D400-Series-Visual-Presets>.
- [5] Tri Khuong Dave Tong John Woodfill Anders Grunnet-Jepsen, John Sweetser. Subpixel linearity improvement for intel realsense depth camera d400 series. <https://dev.intelrealsense.com/docs/white-paper-subpixel-linearity-improvement-for-intel-realsense-depth-cameras>.
- [6] Post-processing filters. <https://dev.intelrealsense.com/docs/post-processing-filters>.
- [7] Guillem Aleny'a Sergi Foix and Carme Torras. Lock-in time-of-flight (tof) cameras: A survey. *IEEE SENSORS JOURNAL*, 11, 03 2011.
- [8] Helios 2 specifications. <https://thinklucid.com/helios-time-of-flight-tof-camera/>.

- [9] Online basler blaze 101 camera documentation. <https://docs.baslerweb.com/blaze-features>.
- [10] Martin Gramatke. High-resolution time-of-flight cameras. 12 2019.
- [11] Basler blaze 101 camera specifications. <https://www.baslerweb.com/en/products/cameras/3d-cameras/basler-blaze/blaze-101/>.
- [12] John Canny. A computational approach to edge detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, PAMI-8:679 – 698, 12 1986.
- [13] Shahrizat Shaik Mohamed, Noorita Tahir, and Rahmi Adnan. Background modelling and background subtraction performance for object detection. pages 1 – 6, 06 2010.
- [14] Yannick Benezeth, Pierre-Marc Jodoin, Bruno Emile, Helene Laurent, and Christophe Rosenberger. Comparative study of background subtraction algorithms. *Journal of Electronic Imaging*, 19:033003–033003, 07 2010.
- [15] Klaus Greff, Andre Brandao, Stephan Kraub, Didier Stricker, and Esteban Clua. A comparison between background subtraction algorithms using a consumer depth camera. *VISAPP 2012 - Proceedings of the International Conference on Computer Vision Theory and Applications*, 1, 02 2012.
- [16] Enrique Fernandez-Sanchez, Javier Diaz, and Eduardo Ros. Background subtraction based on color and depth using active sensors. *Sensors (Basel, Switzerland)*, 13:8895–915, 07 2013.
- [17] Solution guide ii-b matching. <http://download.mvtec.com/halcon-12.0-solution-guide-ii-b-matching.pdf>.
- [18] Solution guide ii-b shape-based matching. <http://download.mvtec.com/halcon-9.0-solution-guide-ii-b-shape-based-matching.pdf>.
- [19] Solution guide ii-d classification. <http://download.mvtec.com/halcon-12.0-solution-guide-ii-d-classification.pdf>.