



INTERNSHIP REPORT CI – 2019 – SPID (ROB)

 <p>ENSTA Bretagne</p>	<p>ENSTA Bretagne 2 rue F. Verny 29806 Brest Cedex 9, France</p> <p>Lefèvre Lucie lucie.lefevre@ensta-bretagne.org</p>	 <p>TNO innovation for life</p>	<p>TNO Oude Waalsdorperweg 63 NL – 2597 AK The Hague</p> <p>Supervisor : Gertjan Burghouts Scientist Gertjan.burghouts@tno.nl</p>
--	--	--	---

Acknowledgement page

I would like to thank Gertjan Burghouts, my supervisor at TNO who gave me part of his time for the project, but also during the months preceding the internship. He sent me several emails to explain the project and give me advices to live in the Netherlands. Then, during the internship, he paid attention to me, and was always available to answer my questions face to face, or by email when he was not absent from the office. He is always enthusiastic about his work and for sharing his knowledge.

I am also grateful to the other TNO members for their provision of expertise in my project, and also for welcoming me well in the team. So, it was a great pleasure for me to work at TNO.

I would like to take the opportunity to express my gratitude to my professor at ENSTA Bretagne and supervisor Luc JAULIN.

Finally, I would not forget to remember Yasmine Najjar, student of ENSTA Bretagne, who did her internship with Mr Burghouts too. She presented me her project in TNO, and advised me to contact Mr Burghouts to apply for an internship too.

Abstract

This report is about my experiences as a TNO engineering assistant intern. During three months from 4th June 2018 till 31th August, I did my internship at TNO, a Dutch organisation for applied scientific research. This internship is a part of my studies at ENSTA Bretagne. During my internship, I was assigned on a project to recognize short-term human actions in video streams. I worked with Louis Valery, another student at ENSTA Bretagne. We were both supervised by Gertjan Burghouts.

I chose to do my internship abroad for several reasons: to validate my cursus, and specially to develop my English and get international experiences. For these reasons, I decided to write my report in English. My application for this internship was supported by Yasmine Najjar, who did her own internship last year with Mr Burghouts. She related her experiences and advised me to apply for an internship. I was very interested in her project, and I contacted directly Mr Burghouts.

This report contains the theory and the project realisation. Some results I have got with Louis are detailed and explained. Throughout this report, I could take a step back on my internship and analyse my work and my behaviour at work.

At the end of my internship, Louis Valery and I orally presented our project in two steps: during fifteen minutes, we presented our work and our results got during our three-month internship, and for about ten minutes, we answered questions of the TNO team. In this report, some figures, and tables are extracted from the PowerPoint presentation, I will precise these documents later in the report.

Contents

Acknowledgement page.....	2
Abstract	3
1. Introduction.....	5
1.1. About TNO	5
1.1.1. History of the company	5
1.1.2. Focus area.....	5
1.2. About the internship	5
1.2.1. Project topic	5
1.2.2. Development of the internship	6
2. Technical descriptions of the project	6
2.1. Tools and software used	6
2.2. Theoretical standpoint	7
2.2.1 Artificial Intelligence (AI)	7
2.2.2 Optical flow estimation	8
2.2.3 Random Forest	9
2.2.5 Support vector machine (SVM)	9
2.2.4 Detection in the project	11
2.3. Detection algorithm structure.....	11
3. Results	17
3.1. Experiments.....	17
3.4.1 Detection of wave	17
3.4.2 Detection of jump and fall.....	19
3.4.3 Detection of wave and fall.....	20
3.2. Comments	20
3.2.1 Limits of the method	20
3.2.2 Possible improvement of the detection	21
4. Assessment.....	22
4.1. Outcome of my internship	22
4.2. Organisation de TNO	22
Conclusion	23
References.....	24
Table des figures.....	24
Annex 1: Organisation chart of the Intelligent Image Group	25

Annex 2: Assessment report 27

1. Introduction

1.1. About TNO

1.1.1. History of the company

TNO (Toegepast Natuurwetenschappelijk Onderzoek means Netherlands Organisation for Applied Scientific Research) is a Dutch independent contract research organisation, focused on applied science with more than 3500 employees. TNO was created in 1932 in order to collaborate with companies and government on innovation and science knowledge. TNO has a complicated status, it is a research organisation and it can operate as an industry. TNO has an annual turnover of 580 million Euros but in the same time, it is a not-for-profit knowledge organisation, therefore all its profits are invested. TNO shares with everybody its annual report, and its financial statements, (e.g. the digital annual report for 2017 [1], and the digital financial statements for 2017 [2]). Its financial statement is issued by the Minister of Education, Culture and Science, that shows strong relationship between TNO and the Dutch government.

1.1.2. Focus area

TNO provides researches, data and innovations for businesses and governments. TNO's purpose is to develop technological advances for practical application in society. Besides, one of the TNO's is "We connect people and knowledge in order to create innovations". It is obviously involved in the Netherlands but also at international scale. TNO focuses its researches in nine main areas:

- Buildings, infrastructure and maritime
- Circular economy and environment
- Defence, safety and security
- Healthy living
- Industry
- Information and communication technologies
- Strategic analysis and policy
- Traffic and transport

More information, and several presentations of many projects led by TNO are returned in the TNO website [3].

1.2. About the internship

1.2.1. Project topic

I was assigned to the Intelligent Imagine research group, part of the area "Defence, safety and security". It is composed of about forty professionals and interns who work together on different projects. In the annex 1, there is an organigram of the team, with their name and their area of expertise. In average, each employee simultaneously works on two or three projects. Obviously, as an intern, I worked on only one project in collaboration with Louis Valery, my supervisor and three other members of the department. This department is specialized in image processing, image

enhancement, image analysis, visual pattern recognition and artificial intelligence. I worked particularly in visual pattern recognition and artificial intelligence.

The main purpose of my project called STARE (Spatio-Temporal Action Recognition Efficient) was to create a system able to recognize short-term human actions in video streams. This system could be used to monitor people in small environments, for example to watch patients in hospital and to detect if they need help (if they are waving or if they fall). Furthermore, human actions can help understanding people, showing their feeling or predicting future actions (For instance, if you are crying, you may hide your face behind your hands). This system must be as accurate and fast as possible. Unfortunately, as you will see in the next parts, these two notions are difficult to combine. Moreover, different problems appeared during my internship according to the behaviours of interest, the people concerned and the scene where the system would be used.

This project adds to the work made by Yasmine last year. Indeed, she worked on a project to detect behaviour using the RGB camera, but this year, TNO wants to work on the depth image of the Kinect 2.0.

1.2.2. Development of the internship

At the beginning of my internship, my supervisor gave me a folder called “stare” with several Matlab scripts, Matlab fonctions and Matlab tools to detect a specific behaviour: the waves. Obviously, these codes were drafts, and were not operational. Detecting human behaviours is a tricky problem because some motions are similar, for instance if you raise your hand to scratch your head, it is the same starting motion as if you are waving. My work was to get around this problem and detect only the interesting motion.

During the first three weeks of my internship, my mission was to discover the project and the team, to understand the algorithm, and to get it operational. Then, I found an optimal system to detect only waves.

When the detection of waves was efficient, my mission was to detect other behaviours. Gertjan let me choose the purpose of my project. It was a self-directed project, but it should keep the line of the initial project.

Every week, two or three members of the Intelligent Imagine group prepare a five minutes presentation about their work and explain it to those who are interested. It was very interesting to discover the different existing projects. At the beginning of August, with Louis Valery, we presented our project to about ten persons. Then, at the end of our internship, we made a final presentation about our work and mainly about our results.

2. Technical descriptions of the project

2.1. Tools and software used

Before my internship, during my first exchanges with my supervisor, Gertjan informed me that I would have to bring my own computer, with Windows-64 and Matlab. During the internship, I only used Matlab. Yet, in my office, I had access to a more powerful computer. This computer was necessary to record videos, because my computer missed lot of frames. Furthermore, some computations were too cumbersome, and only TNO’s computer could do it. In addition, TNO issued a Kinect 2.0. I used this camera from the third week to record new videos, and to test the different models on live.

Kinect 2.0 is a sensor which records RGB images and depth ones. In my project, only depth images are used to detect human behaviours in a video.

Depth image is an image in which each pixel of the image relates to the distance from the Kinect to seen objects.

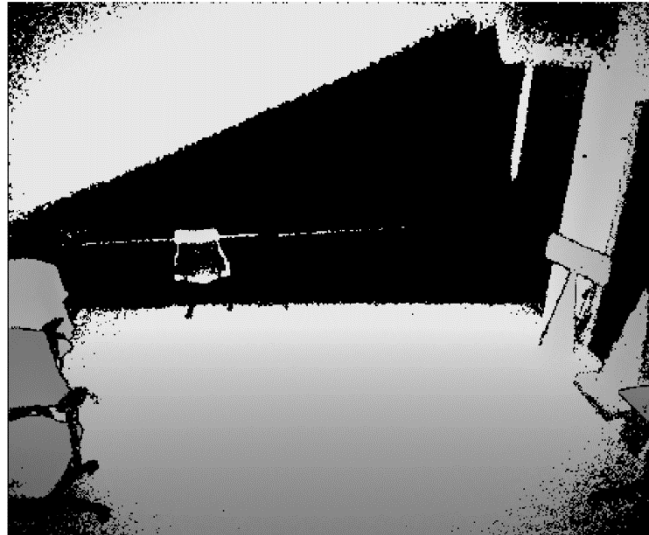


Figure 1 : Depth image from a Kinect 2

The Kinect 2 has a depth image resolution of 512 x 424 pixels and it runs at 30 frames per seconds (FPS). Moreover, the Kinect 2 returns unsigned 16-bit integer frames. However, the algorithm only uses uint8 image. Therefore, I had to initially convert all the depth images in unsigned 8-bit integer images. Furthermore, the Kinect 2 acquires accurate depth value for an object located less than 4500mm from the sensor, other values are considered as noise. In the program, I had then to put at 0 values higher than 4500. And, last point, in the program, the values of the pixel have to be between 0 and 255.

In a depth image, the lightest areas reflect the farthest objects, and inversely, the darkest areas correspond to the closest objects. Yet, be careful, black pixels (value of 0) correspond to either areas located more than 4500mm or noises.

2.2. Theoretical standpoint

2.2.1 Artificial Intelligence (AI)

For this project, an artificial intelligence, and more particularly machine learning is used to detect human behaviours in a video. The machine learning goal is to create a system able to get answers from data. In my project, the answer is to know if a behaviour in a video is a wave or not. This question answering system that we are building is called a **model**. If several answers are possible, then few models will be created, one for each answer. Lots of models are created, and tested, and the final goal is to find the best model. The process to create the model is called **training**. In order to train a model, several data are needed. In my case, data are collected from videos with a person walking, moving, and sometimes waving. Models are training on these videos.

To collect data, we need to choose which factors are efficient to answer our question. The more data are used, the more accurate is the model. These factors are called **features**. In order to detect waves in a video, a relevant factor to recognize behaviour is the motion between successive frames. Indeed, an optical flow estimation enables to describe motion in a video. This step of data collecting

is called **Gathered Data**. In this step, data have to be annotated thanks to its features. For instance, for the detection of wave, data are the intervals collected from the video, in each interval, the motion is known and then each interval is annotated with “other” or “wave”.

Then, data have to be loaded and prepared, and sometimes adjusted for the machine learning training, it is called the **Data preparation**. In this step, the data are gathered and cluttered. The goal of this step is to prevent the system from memorizing the solution. The dataset is separated in two parts, one part is used in training the model, and the other part allows to evaluate the performance of the trained model. Data should not be used for the training and for the evaluation. For the detection on Kinect video, depth images need to be converted.

The next step is to choose a model. To detect human behaviour, two options were possible: Random Forest (RF) or the Gaussian mixture modelling (MoG). These methods quantify the features into patterns.

When this step is over, it is time for the **training**. During the training, the process uses the model chosen in the previous step and tries it with random values. Knowing some outputs, the model prediction can be compared with the outputs Data are here used to incrementally improve the model’s ability to predict if the motion is a wave or not. Several tests are made on different dataset, and after a lot of tests and corrections, the model is improving.

Then, once training is finished, the model created is assessed, this step is called **evaluation**. The model is now tested on the part of the dataset that had never been used during the training step.

2.2.2 Optical flow estimation

Optical flow estimation is a method that shows the image is changing due to motion. For that, it computes to each pixel of the current image, two-components velocity vector indicating the position of the same pixel in the reference image. Assuming that the displacements between these two images are small, two-components are calculated with the local Taylor series approximations, using partial derivatives : $image1 = image2(x + \Delta) = Image2(x) + \frac{\partial Image2}{\partial x} * \Delta x + \frac{\partial Image2}{\partial y} * \Delta y$, with $\Delta = (\Delta x, \Delta y)^T$. With this equation, the displacement at one pixel is estimated.

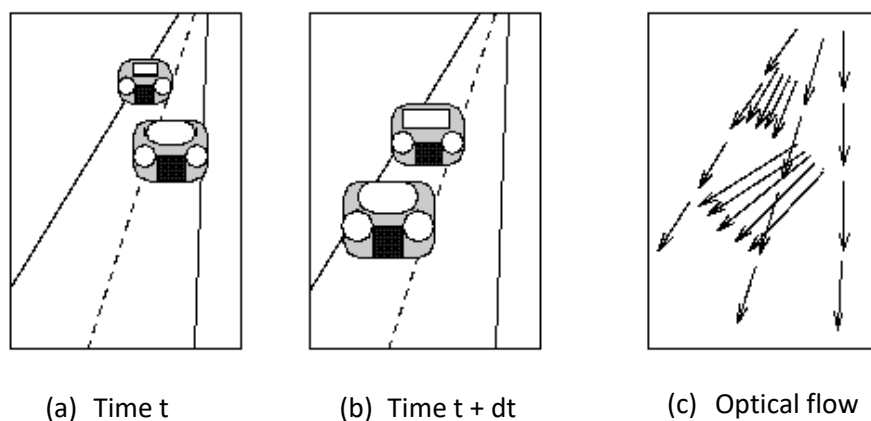


Figure 2 : Illustration of the Optical flow

2.2.3 Random Forest

Random forest is an algorithm used for classification and regression tasks. In machine learning, RF enable to have an accurate and stable prediction, choosing adequate parameters. To understand Random Forest, Decision Trees must be known.

Decision Tree is a map of different possible results or answers for an experiment or a question with different parameters. Each branch of the tree is browsed through. Decision Tree can find irrelevant parameters, and it creates some set of rules. Then, the rules can be used to find the answer or to predict the results of an experiment. For example, if you want to buy a new book, but you do not know which one buys. Therefore, you will ask several friends what kind of books they read, and if they like them. With all the answers, you will have advice to choose your next book.

A Random Forest regroups several Decision Trees. The larger number of trees, the more accurate the results will be. It searches for the best feature among a **random** subset of features. Therefore, when non-relevant results are found, the score combined with this feature will decrease. Thus, Random Forest can classify features. It enables to improve the prediction of the models thanks to parameters chosen for the best feature. For instance, with the previous example, this time, you will start asking more and more advises from your friends and family, and they will give you some recommendations. With all the answers, and the recommendations, you will choose the book which is the most recommended.

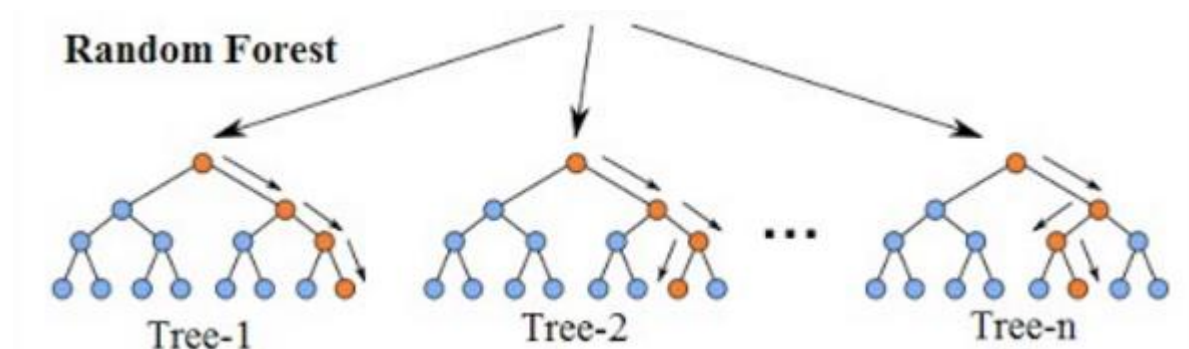


Figure 3: Random Forest

2.2.5 Support vector machine (SVM)

Support vector machine is another model used in machine learning. Like Random Forest model, it can analyse data for classification and regression. SVM builds a model in which data are split in two categories. The model is represented as a map with the different points located in each category. These two categories are separated by a gap which is as small as possible. The output of the algorithm is an optimal hyperplane representing the gap.

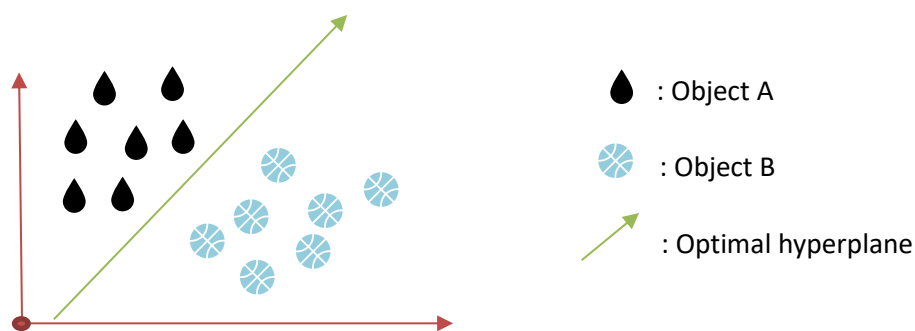


Figure 4: SVM applied on data with only two types of object

Yet, in general, two categories are rarely separated by an affine line. To solve this problem, the algorithm has different solutions:

- Regularization

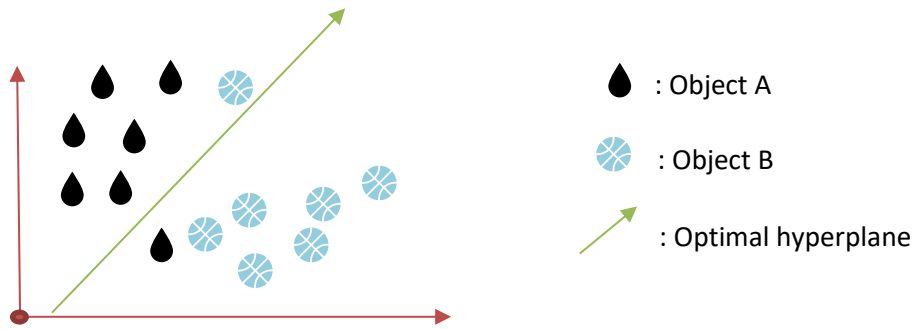


Figure 5: SVM using regularization

- Transformation of the map (Kernel)

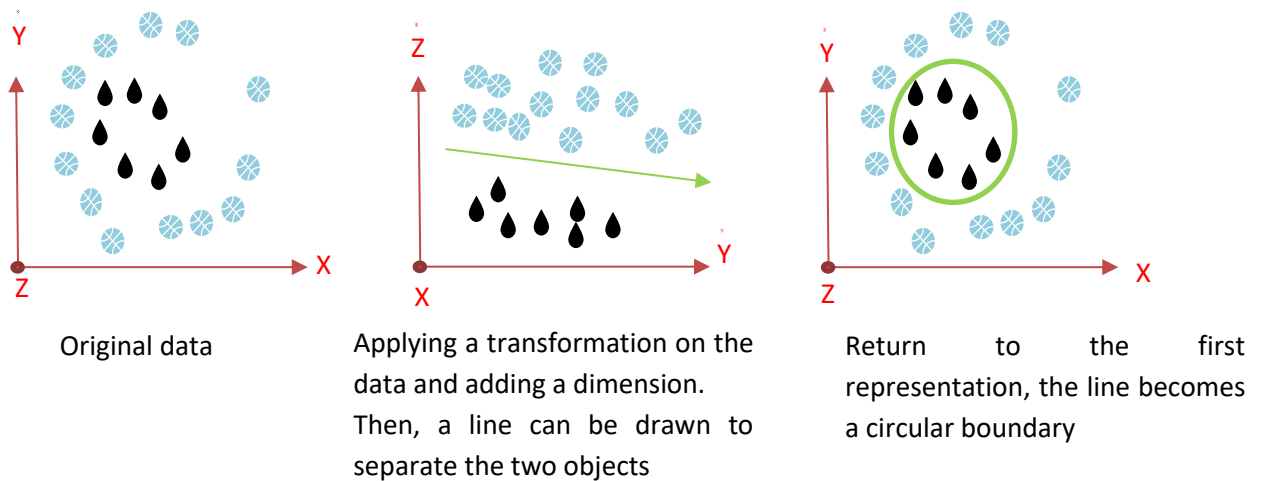


Figure 6: SVM using a Kernel

2.2.4 Detection in the project

In this part, I will outline the detection mechanism in my project, stating the main components, to a better understanding of the next chapter.

The detection is made on spatio-temporal volumes, those containing a motion are called bounding boxes (bboxes). Bboxes are defined by their width, height and temporal duration (called history seconds).

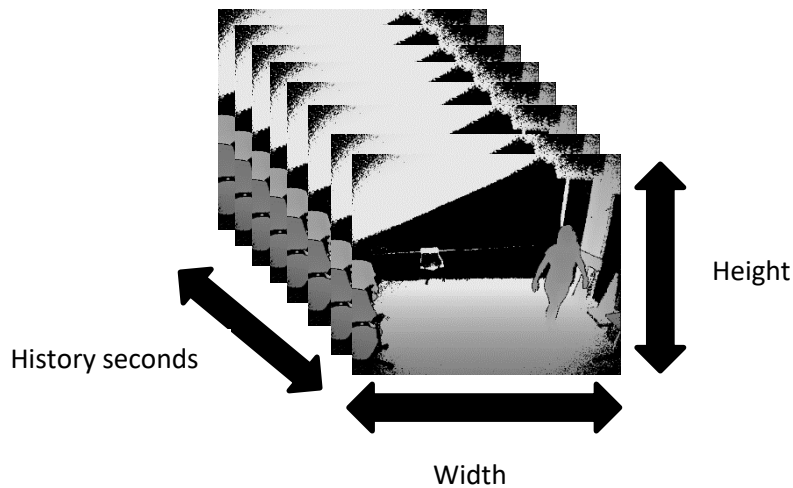


Figure 7: Representation of a bbox

2.3. Detection algorithm structure

In the part 2.2.1, the method of machine learning has been explained. In this part, the structure of the detection of wave using machine learning is to be described.

In this project, the detection of wave uses each step described in the 2.2.1 Artificial Intelligence (AI). I recorded several videos lasting around three or four minutes. These videos are all recorded in the same room, with the same view angle. In each video, a person is walking, waving, jumping, etc, in the whole place of the scene. It is possible to have several persons on the same video, in order to have an accurate model, trained in different situations.

During the **Gathered Data**, one of the recorded videos is chosen, with the behaviours related to the project. This video is split in several intervals, these intervals constitutes the **dataset**. This dataset will be used to train the model. Each interval is annotated, that means its behaviour is specified with an action label (“wave” or “other” for the intervals without the wanted behaviour).

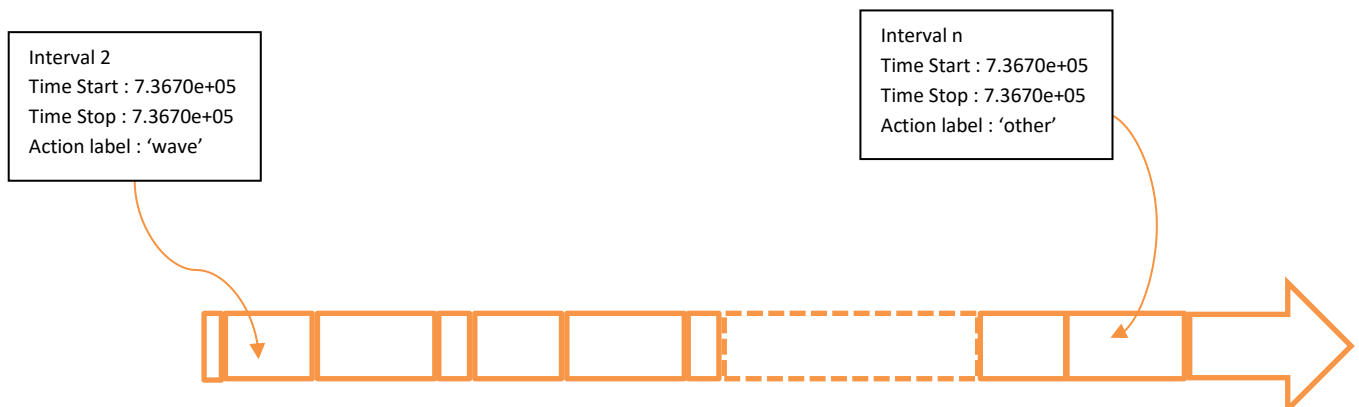


Figure 8: Splitting of the video in several intervals

The annotation also corresponds to the creation of bounding boxes (called bboxes) around the action. Each frame of the interval “wave” is annotated manually, that means a rectangle is drawn around the motion, step shown in Figure 9 : Annotation of a wave. Each action occurrence is designated as one **event**, it is a set of frames.

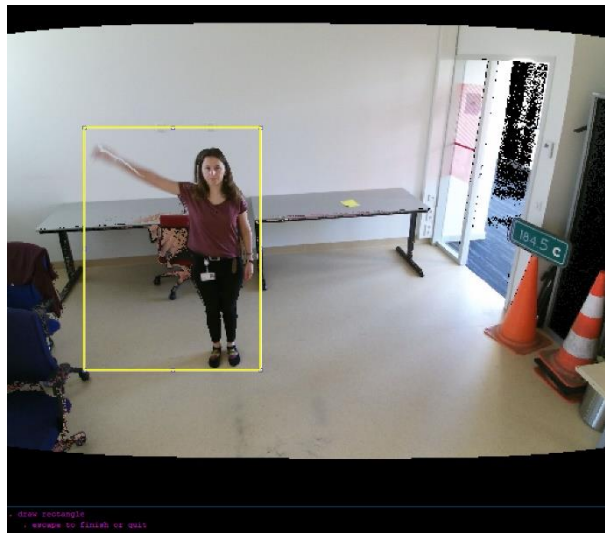


Figure 9 : Annotation of a wave

This supervision is required to localize actions in the frame because, there are lot of different actions in a video (noises, winds, person walking, etc...), a human being often performs several actions, but only waving behaviours are significant for the detection.

During the **Data preparation**, human supervision is needed to indicate:

For the scene only:

- The size of human in back stage
- The size of human in front stage
- The area of interest (where humans can be), for example in order to detect wave, the ground doesn't need to be checked

These three components make the calibration of the scene.

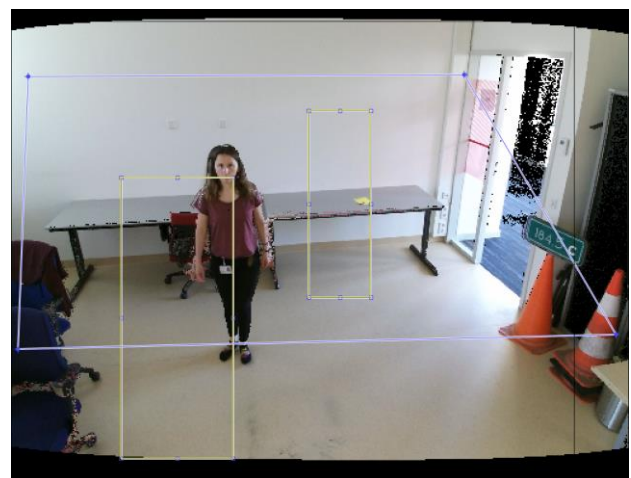


Figure 10: Calibration of the scene to detect a wave, in blue the area of interest, in yellow the size of a human being in the front and in the back

For the actions:

- Action labels: “wave” (if you want to detect other behaviours, it is where you have to add the action label corresponding, for example “jump”)
- Height of the spatio-temporal volume (in meter)
- Width of the spatio-temporal volume (in meter)

These parameters define the size of the different bboxes on the scene.

A grid is then created for each interval, it is a representation of the bboxes that might be contained in an interval. For that reason, the choice of parameters is crucial, because the number of bboxes and their size depends on these parameters among other ones. Furthermore, the grids calculate the position of the points in the bboxes, where the optical flow will be computed. In Figure 8, an example of grids is represented. Moreover, grids are useful to see how the detector will detect wave in the video. Several times during my project, I checked the grids and changed the parameters of the bboxes depending on the grids. For example, in the grids of Figure 8, bboxes in the back stage (in black) might be too small to detect a wave, therefore I increased the size of the height, and tested again the model.

*

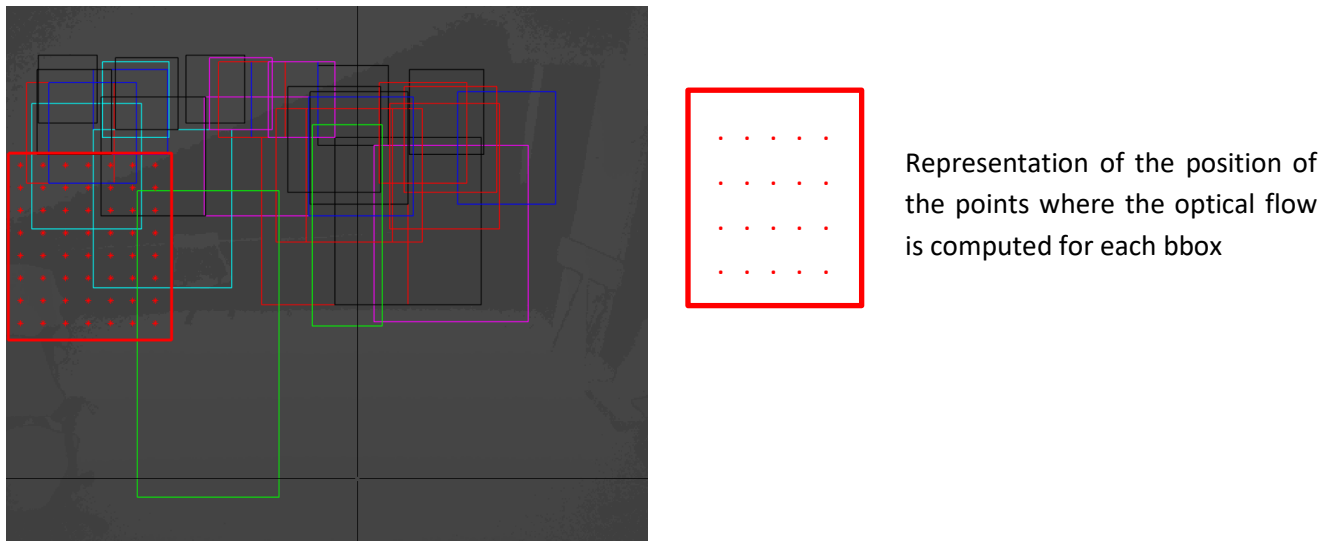
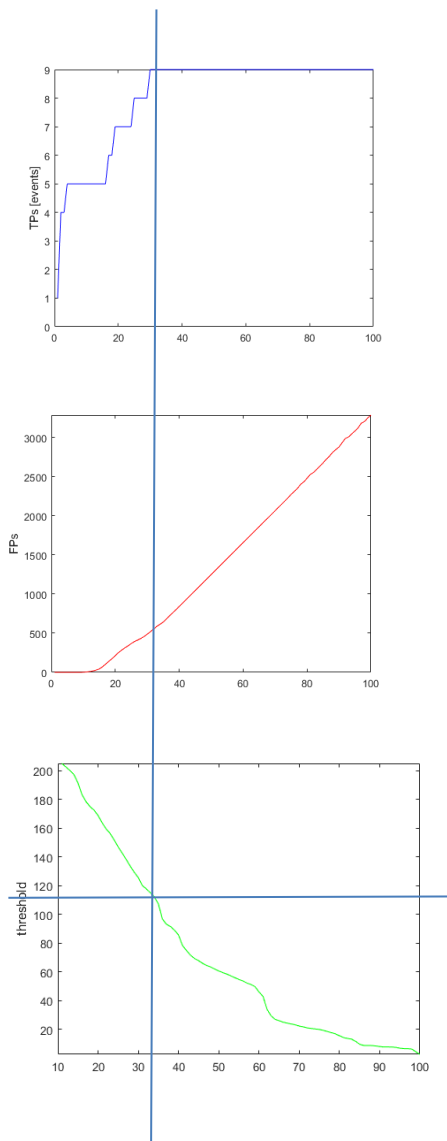


Figure 11 : Example of Grids

During the **training mode**, the first processing step called **detector** is to select bboxes that will be used in the second processing step. This step allows to decrease the computing time, reducing the number of relevant bboxes. The factor of selection is the number of motions in a bbox, indeed if there are enough motions, that may include an action is happening. In each bbox, a detector score is calculated and quantifies the motions in this bbox. At the same time, the bboxes are compared with the bboxes annotated manually during Gathered Data. Only bboxes which overlap bboxes annotated are kept.

Then, a workpoint threshold is chosen (automatically or manually), taking into account the number of events detected and the number of “false positive” detected. The false positive corresponds to behaviours which are detected as wave, but that are not. For instance, if you raise your hand, it can be detected as a wave. In this case, this behaviour is called “false positive”. To choose manually the threshold, three graphs are useful. In Figure 12: Choice of the threshold for the detector ,these graphs are shown, and the method of choice is explained. The bboxes with a detector score higher than the threshold remain for the rest of the process.



The first graph shows the number of the event corresponding to the wave detected. In this example, the number of waves in the video is 9. Therefore, the number of events should be 9.

The second graph shows the number of the false positive detected. This number should be as small as possible.

In this graph, the threshold, corresponding to the choice made with the two other graphics, is saved. In this example, the value of the threshold is about 115.

Figure 12: Choice of the threshold for the detector

After this step, only the bboxes with a sufficient value of motion are kept.

The second processing step is to compute features within the remaining bboxes chosen in the detector. The grid gives the position of the point in the bbox, where the optical flow estimation computes the features. Then, an action is predicted for each bbox.

Before these steps, a human supervision is also needed to choose processing parameters:

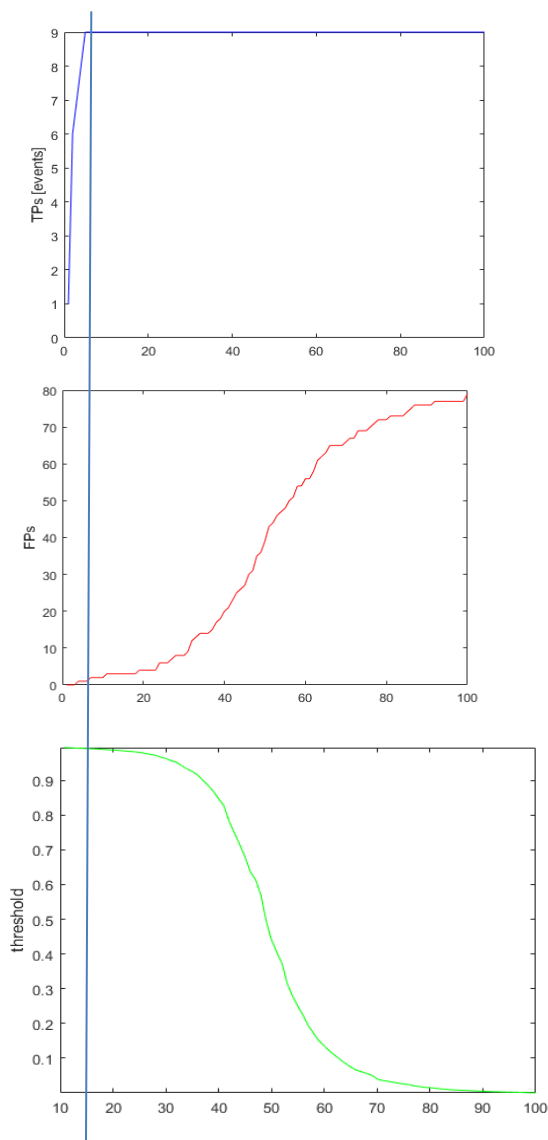
- The number of features per bboxes
- The resize factor for the bboxes computation
- The resize factor for the computation of optical flow within a bbox
- The temporal duration of a bbox

Once these steps are over, a model can be created.

The following parameters must be indicated by human supervision:

- Choice of the **quantifier** and the **classifier**, in my project, it is Random Forest method and Support Vector Machine method
- Number of quantifications indicates the patterns in which the features are quantified
- Number of samples to train the classifier

During this new step, first, the features computed with optical flow in the remaining bboxes are quantified into patterns with the RF method (a mixture-of-Gaussians could have been used). The RF method gets a statistical description of the optical flow. Then, the features are classified with SVM method. This method searches the features which correspond to a wave, otherwise, the features are classified in the category "other". Then, a new threshold is chosen. It selects bboxes corresponding to the desired behaviours. Indeed, each bbox is classified and has a percentage of wave. If its percentage is higher than the threshold, therefore the bbox becomes "positive".



The first graph shows the number of events corresponding to the wave detected after the classifier. In the video, there are 9 waves. Therefore, the number of positive events should be 9 after the classifier.

The second graph shows the number of the false positives detected after the classifier. This number should be as small as possible.

In this graph, the threshold, corresponding to the choice made with the two other graphics, is saved. In this example, the value of the threshold is 1. We always chose a value lower than 1, for example 0.99, because when we test the model in live, the detection is never perfect.

Figure 13: Choice of the threshold for the classification

Once created, the model can be tested on the video wanted. Two functions are available: one predicts behaviours on replay, with any recorded videos, the other predicts behaviours on live with the Kinect 2.0. In these two cases, the viewing angles and the scene must be identical to the training video.

To improve our model, different parameters can be changed and tested to figure out the good parameters. The performance of the model is assessed. A function named "Optimization parameters" computes two values `mp_all` and `mp_best` for different parameters (size of the bboxes, history second, number of quantification and number of features). `mp_all` and `mp_best` assess the detection. Optimization parameters function is a powerful tool to optimize a model with one behaviour. However, when few behaviours need to be detected, optimization parameters function is inefficient. Indeed, `mp_all` and `mp_best` are falsed by the different behaviours; during the detector, if one action is not detected, `mp_all` do not take account of it.

3. Results

3.1. Experiments

3.4.1 Detection of wave

At the beginning, I worked on a three-minutes recorded video. During the video, I am waving nine times and doing other behaviours in the rest of the video. The purpose of this first mission was to detect wave.

A good model with a good detection of wave on recorded video and on live video are made with the following parameters:

- Bbox height: 1.6 meter
- Bbox width: 1.4 meter
- History seconds: 2 seconds
- Number of quantifications: 64
- Number of feats: 100
- Classifier threshold: 0.99

These parameters have been found after several tests, and after using an optimization parameters function.

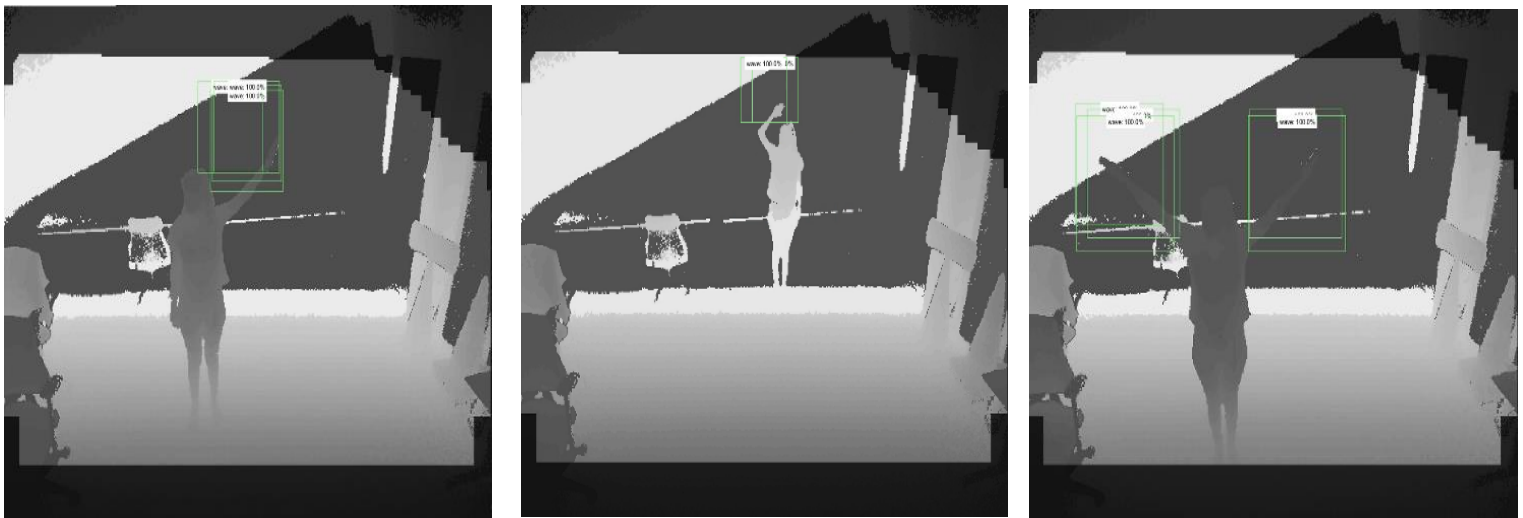


Figure 14: Detection of wave in different place in the scene

The green bboxes correspond to a detected wave, the classifier threshold is worth 0.99, therefore, bboxes with a percentage of wave higher than 99% are detected as wave (in green). Other bboxes appear when motions are detected, but the bboxes related have a percentage of motion lower than 99%, so this motion is not detected. In the Figure 15: Bbox corresponding to an "other" motion, the actor is walking, it is the reason why the percentage of waving is about 0.4%.

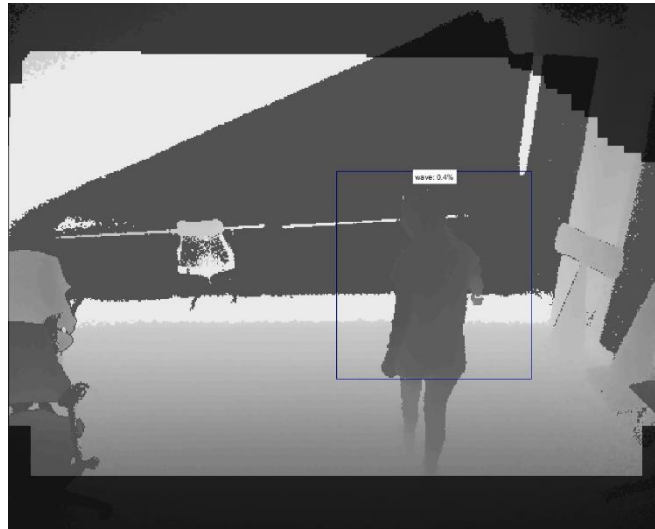


Figure 15: Bbox corresponding to an "other" motion

This model is able to detect two persons waving at the same time in the video. Afterwards, the next step was to detect two different behaviours in a same video.

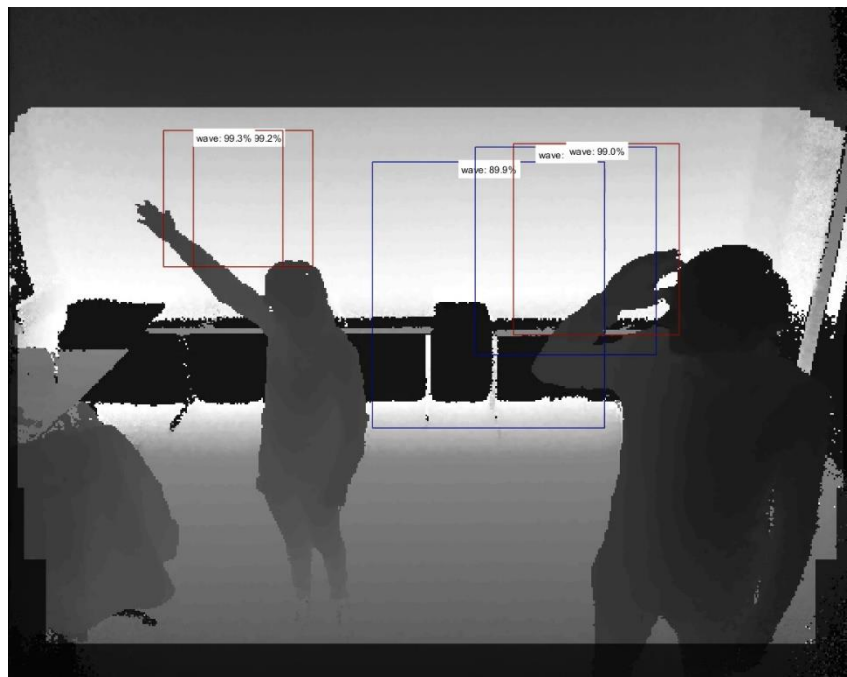


Figure 16: Two persons waving at the same time

To have a better detection of two persons waving at the same time, it is better to have this example in the dataset used to obtain the model.

Furthermore, to detect two persons waving at the same time, the threshold for the classification must be lower. Indeed, in average, for each bbox the percentage of waving is around 92% when two waves are detected, whereas when one person is waving the average percentage for a positive wave is 98%.

The table in Figure 15 illustrates my points. It represents the percentage of bboxes appearing, depending on the present behaviours. Indeed, when we are two in front of the Kinect 2, more “other” bboxes appear.

	Wave 1m60	Wave 1m80	Other
Wave	41	43	16
Other	0	0	100



Figure 17: Table of the results for the detection of two people waving

3.4.2 Detection of jump and fall

P1	Wave	Jump	Other	P2	Wave	Jump	Other
Wave	95	4	1	Wave	96	4	2
Jump	3	91	9	Jump	2	92	7
Other	2	5	90	Other	2	4	93

P3	Wave	Jump	Other
Wave	95	3	1
Jump	4	89	3
Other	1	8	96

P4	Wave	Jump	Other	P5	Wave	Jump	Other
Wave	95	17	5	Wave	95	16	6
Jump	2	79	6	Jump	0	79	7
Other	3	4	89	Other	0	5	87

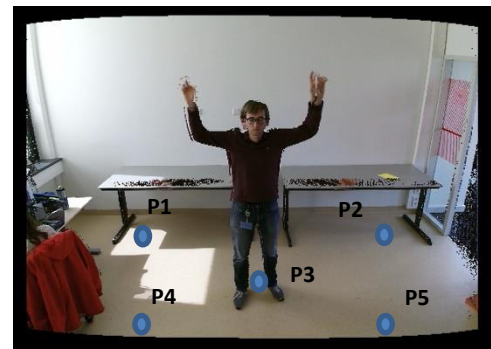


Figure 18: Table of the results for the detection of wave and jump

In general, wave and jump are correctly detected by the model, but some errors remain. The table of the Figure 16, the percentage of which bboxes appearing depending on the present behaviours. At the front part of the scene (P4 and P5) some waves are detected as a jump. In the middle of the scene (P3) some other behaviours are detected as a jump. And finally, at the back part of the scene (P1 and P2), some jumps are not detected.

3.4.3 Detection of wave and fall

One difficulty met was to detect two behaviours in the same video with one model.

The next step of my internship was to detect more behaviours with one model. I decided to detect fall, indeed the detection of fall can be useful in indoor room with patients who need to be watched.

I used a new recorded video to train the model, with 11 waves, and 9 falls. To detect wave, a history seconds of 2 seconds was optimal, but falls cannot be detected with this value, as a fall action lasts less than 2 seconds. After several tests, I found how to detect fall, a history second of 0.15 seconds was needed.

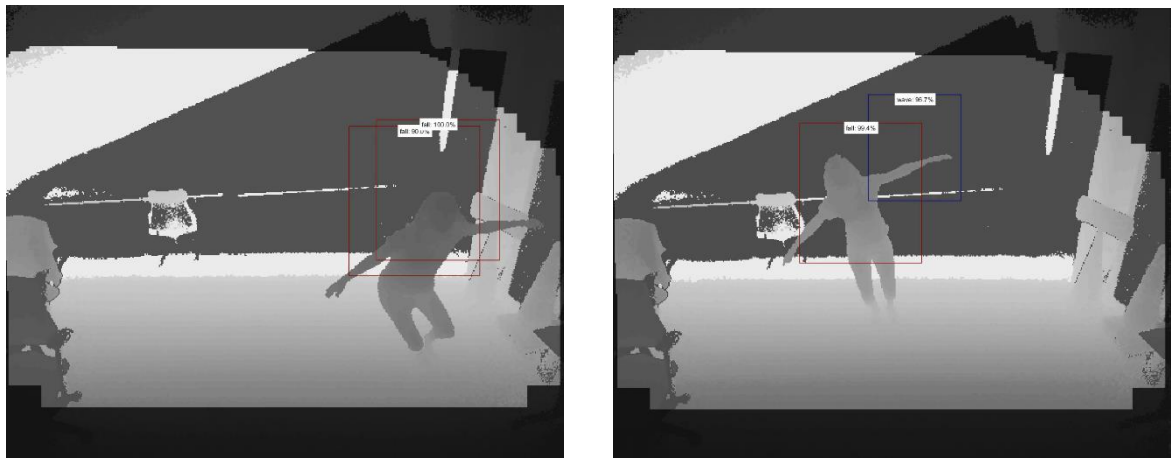


Figure 19: Detection of fall

Adding two behaviours outlines the difficulties encountered by the process to distinguish two similar behaviours. In Figure 19: Detection of fall, the picture at the right, a wave might be detected during the fall.

3.2. Comments

3.2.1 Limits of the method

The process of detecting short-term human actions meets different difficulties.

- Calibration

The detection of actions in the front or in the back parts of the scene is more difficult, the calibration of the scene has to be very accurate.

In some cases, with Louis, we trained models on different videos, and with the same model. In the front, Louis waving was correctly detected whereas, at the same place, I was not detected.
- Similar behaviours

When other behaviours look like behaviour of interest, false positives appear.

-Raising your hand vs waving

-Walking vs jumping

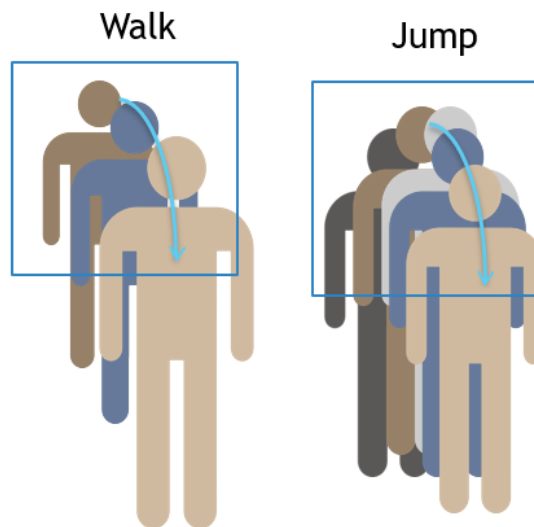


Figure 20: Explanation of the false positive between walking and jumping

In this figure, the similarity of both motions; walking straight perpendicularly to the camera and jumping is underscored. Therefore, these two bboxes are classified in the same event “jump”.

- Size of bbox

To detect different behaviours with one model, if the behaviours are too different, it is hard to find the correct bbox to detect all the behaviours

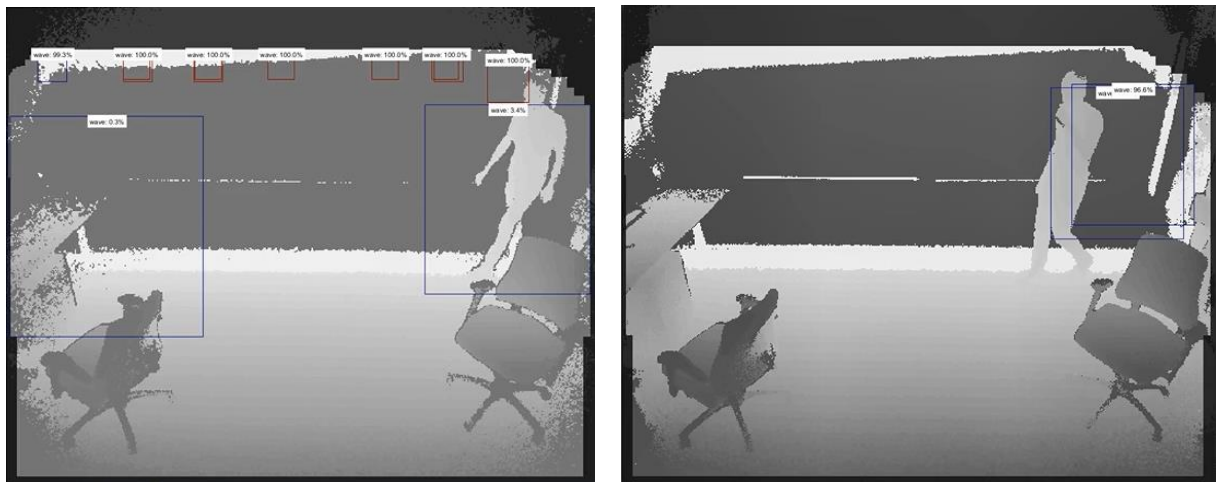
For example, to detect waves, the optimal history seconds is around 2 seconds, whereas to detect falls, the history seconds cannot exceed 0.3 seconds.

Similarly, to detect jumps the width of bbox is around 0.7m, whereas to detect waves, the optimal width of bbox is 1.4m.

3.2.2 Possible improvement of the detection

A set of enhancements can be accomplished on the detection of short-term human actions. One possibility would be to improve the computing speed. In the program, timestep_seconds is a value corresponding to the time after which the next frame is processed. To detect short-actions as jump or fall, all frames need to be processed, but to detect wave, timestep_seconds can be worth 0.3 or 0.4. Timestep_seconds cannot be higher because it influences on the accuracy of the model.

A second possibility is to delete noises in the depth image. Undeniably, noises are interfering in the detection; indeed, noises can be detected by the process as a motion. To delete noise, a mask is used for each frame. In Figure 21: Deletion of noises in the right image, few bboxes appear in the top of the scene, these bboxes are false positives.



(a) Frame directly recorded by the Kinect 2

(b) Frame directly recorded by the Kinect 2

Figure 21: Deletion of noises

4. Assessment

4.1. Outcome of my internship

During my internship at TNO, I have obviously acquired knowledge on artificial intelligence, especially machine learning, and on image processing. These processes were quite unclear for me before my internship, because I had no background on these activities. Yet, Gertjan explained me clearly this notion. The first day, during more than one hour, he gave me the process in detail, and how to use the different functions. He advised me to read some studies made by TNO on the realm. Then, I learned the job. Furthermore, before this internship, I had a basic knowledge of the programming Matlab language, but for three months I exclusively used it, thus, now, I am comfortable with it.

Throughout my internship, I have met many people expert in different sciences (mathematics, physics, management, etc), it was very interesting, and it was a very good opportunity to exchange with them on different topics.

Finally, this internship was an instructive international experience for me. At TNO, but also during the remaining of my time in the Netherlands, I had the chance to exchange with Dutch people, and other foreign ones, about work but also about their culture and hobbies. I practiced spoken English as well as written English. I have increased my vocabulary, and I have a better understanding even if my accent has not improved a lot.

4.2. Organisation de TNO

In the future, I would like to work in robotics. It was not the main topic of my internship, but I think it is important to know different approaches of technological subjects. Moreover, machine learning can be applied to robotics, an example of it is described by Evan Ackerman in his article [4]. This robot called

Watch-Bot is processed to learn the household activities of its owner and detect when something is wrong. Furthermore, there were several other robotics students who did an internship or their PhD (Philosophiae doctor) at TNO. I could discuss with them about their knowledge and their projects.

At the end of my internship, Gertjan filled the assessment report, and we had a short meeting together to discuss about it, and to get my feedback about my internship. It was very interesting, because I know how I can improve my behaviours in my future career. I should not hesitate to ask more questions, and to be more communicative about my future projects' progress.

Finally, the last day of my internship, I made a presentation about my project with Louis. It has been a good practice for our future work, at the ENSTA, or in our future career.

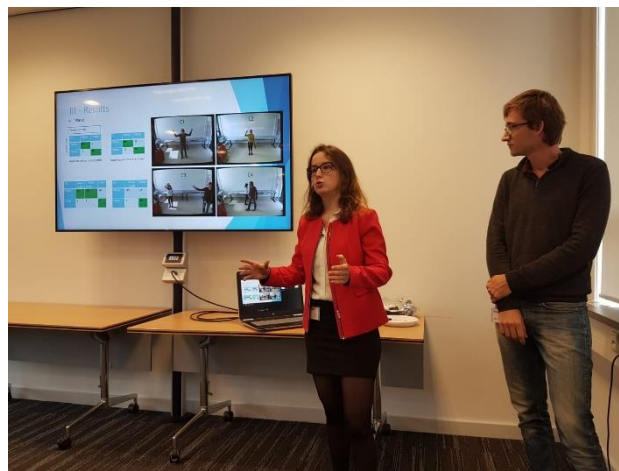


Figure 22: Oral presentation, Tuesday 28th August 2018

Conclusion

My project main purpose was to study the efficiency of the Kinect 2.0 in order to recognize short-term human actions. A major advantage of the Kinect 2.0 is its capacity to collect RGB-image and Depth-image. After my researches, and my experiments described in my report, I can conclude that the Kinect 2 using the Depth image is able to detect different behaviours in a video, yet some limitations of the problem could be solved with the RGB-image. Therefore, the Kinect 2.0 is an efficient sensor to detect specific human behaviours. A future work on this project could be a mixture of systems of detection using these two data.

This internship has been a rewarding experience, I feel I have learned a lot and I became more autonomous.

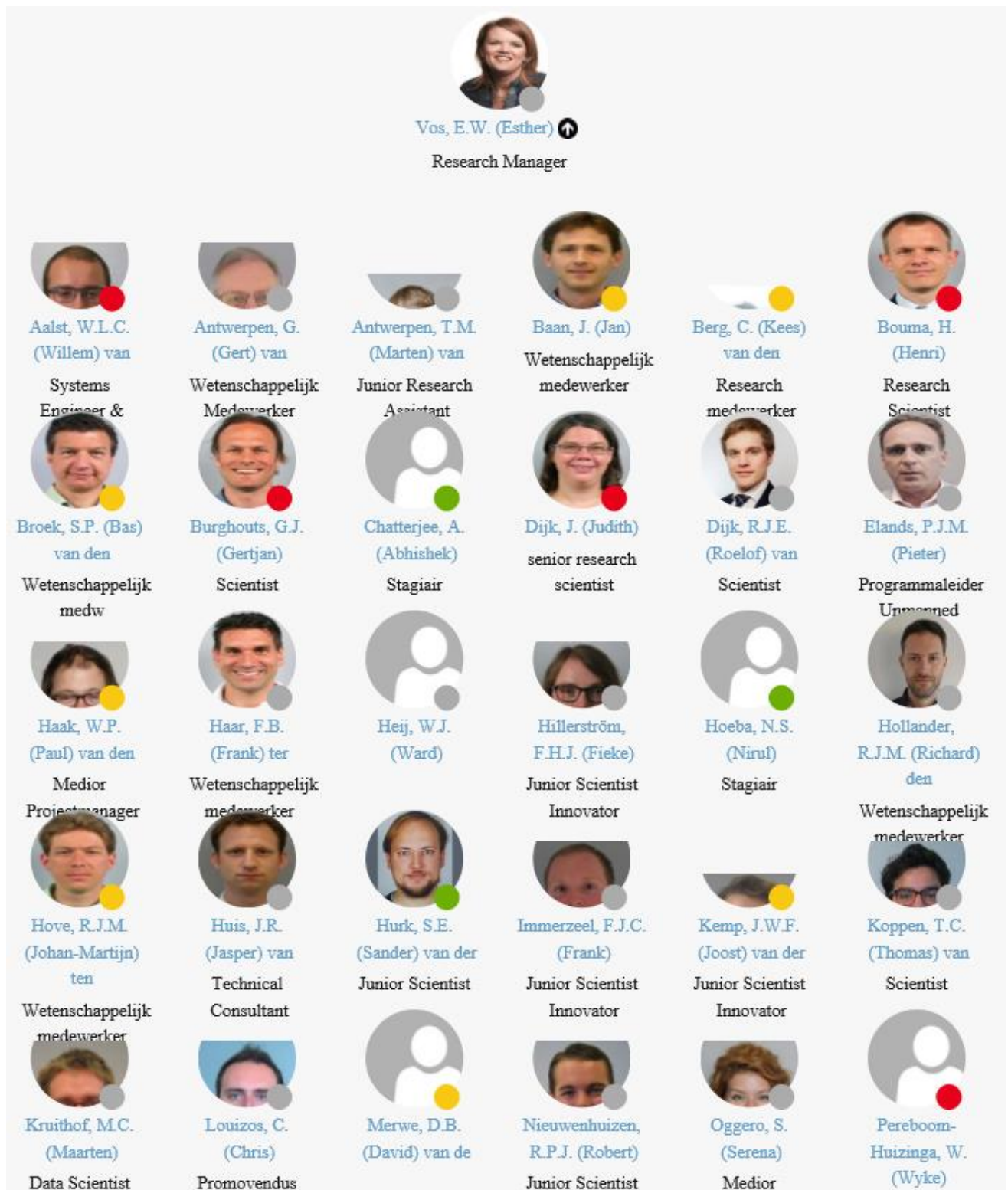
References

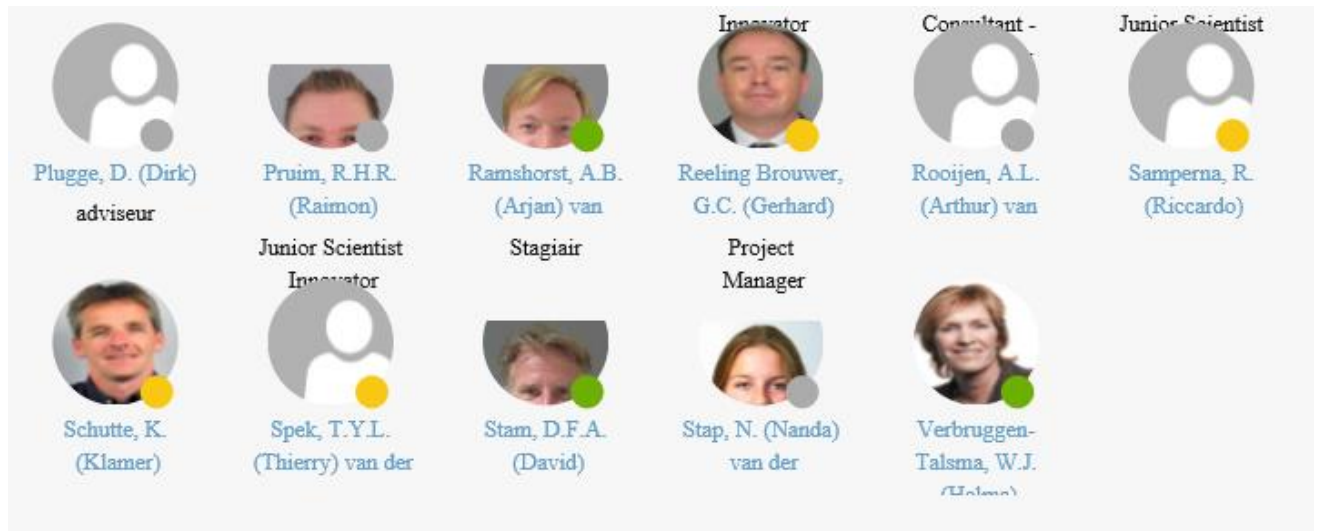
- [1] <https://view.publitas.com/cfreport/tno-jv2017-annual-report/page/1>
- [2] <https://view.publitas.com/cfreport/tno-jv2017-financial-statements/page/1>
- [3] TNO – innovation for life. <https://www.tno.nl/en/>
- [4] Daniel Faggella (September 16th 2018). Machine Learning in Robotics – 5 Modern Applications. Retrieved from <https://www.techemergence.com/machine-learning-in-robotics/>

Table des figures

Figure 1 : Depth image from a Kinect 2.....	7
Figure 2 : Illustation of the Optical flow	8
Figure 3: Random Forest	9
Figure 4: SVM applied on data with only two types of object	9
Figure 5: Representation of a bbox.....	11
Figure 6: Splitting of the video in several intervals.....	11
Figure 7 : Annotation of a wave	12
Figure 8: Calibration of the scene to detect a wave, in blue the area of interest, in yellow the size of a human in the front and in the back.....	12
Figure 9 : Example of Grids.....	13
Figure 10: Choice of the threshold for the detector	14
Figure 11: Choice of the threshold for the classification	15
Figure 12: Detection of wave in different place in the scene	17
Figure 13: Bboxes corresponding to an "other" motion	18
Figure 14: Two people waving in the same time	18
Figure 15: Detection of fall.....	20
Figure 16: Deletion of noises.....	22

Annex 1: Organisation chart of the Intelligent Image Group





Annex 2: Assessment report



RAPPORT D'EVALUATION ASSESSMENT REPORT

Merci de retourner ce rapport en fin du stage à :
Please return this report at the end of the internship to :

ENSTA Bretagne – Bureau des stages - 3 rue François Verny - 29806 BREST cedex 9 – FRANCE
☎ 00.33 (0) 2.98.34.87.70 - Fax 00.33 (0) 2.98.34.87.90 - stages@ensta-bretagne.fr

I - ORGANISME / HOST ORGANISATION

NOM / Name TNO

Adresse / Address Oude Waalsdorperweg 62

Tél / Phone (including country and area code) +31 648842445

Fax / Fax (including country and area code) _____

Nom du superviseur / Name of placement supervisor Gertjan Burghouts, PhD

Fonction / Function senior researcher

Adresse e-mail / E-mail address gertjan.burghouts@tno.nl

Nom du stagiaire accueilli / Name of trainee Lucie Lefevre

II - EVALUATION / ASSESSMENT

Veuillez attribuer une note, en encadrant la lettre appropriée, pour chacune des caractéristiques suivantes. Cette note devra se situer entre A (très bien) et F (très faible)
Please attribute a mark from A (very good) to F (very weak).

MISSION / TASK

❖ La mission de départ a-t-elle été remplie ? A B C D E F
Was the initial contract carried out to your satisfaction?

❖ Manquait-il au stagiaire des connaissances ? oui/yes non/no
Was the trainee lacking skills?

Si oui, lesquelles ? / If so, which skills? video processing, video analysis,

ESPRIT D'EQUIPE / TEAM SPIRIT

❖ Le stagiaire s'est-il bien intégré dans l'organisme d'accueil (disponible, sérieux, s'est adapté au travail en groupe) ? / Did the trainee easily integrate the host organisation? (flexible, conscientious, adapted to team work) A B C D E F

Souhaitez-vous nous faire part d'observations ou suggestions ? / If you wish to comment or make a suggestion, please do so here

Lucie is very nice and smart, but can be more assertive in asking for help.

Versión du 03/05/2017

COMPORTEMENT AU TRAVAIL / BEHAVIOUR TOWARDS WORK

Le comportement du stagiaire était-il conforme à vos attentes (Ponctuel, ordonné, respectueux, soucieux de participer et d'acquérir de nouvelles connaissances) ?

Did the trainee live up to expectations? (Punctual, methodical, responsive to management instructions, attentive to quality, concerned with acquiring new skills)?

A B **C** D E F

Souhaitez vous nous faire part d'observations ou suggestions ? / *If you wish to comment or make a suggestion, please do so here*

Lucie can be more communicative about progress, status and problems. To some suggestions, I did not get a confirmation or response.

INITIATIVE – AUTONOMIE / INITIATIVE – AUTONOMY

Le stagiaire s'est-il rapidement adapté à de nouvelles situations ?

(Proposition de solutions aux problèmes rencontrés, autonomie dans le travail, etc.)

A B C D E F

Did the trainee adapt well to new situations?

(*eg. suggested solutions to problems encountered, demonstrated autonomy in his/her job, etc.*)

A B **C** D E F

Souhaitez-vous nous faire part d'observations ou suggestions ? / *If you wish to comment or make a suggestion, please do so here*

Lucie very well developed a demo and performed experiments for various, new questions that arised

CULTUREL – COMMUNICATION / CULTURAL – COMMUNICATION

Le stagiaire était-il ouvert, d'une manière générale, à la communication ?

Was the trainee open to listening and expressing himself/herself?

A B **C** D E F

Souhaitez-vous nous faire part d'observations ou suggestions ? / *If you wish to comment or make a suggestion, please do so here*

very keen to learn, improved the question-asking during the internship, got more assertive and autonomous

OPINION GLOBALE / OVERALL ASSESSMENT

❖ La valeur technique du stagiaire était :

Evaluate the technical skills of the trainee:

A B **C** D E F

III - PARTENARIAT FUTUR / FUTURE PARTNERSHIP

❖ Êtes vous prêt à accueillir un autre stagiaire l'an prochain ?

Would you be willing to host another trainee next year? oui/yes non/no

Fait à _____, le _____
 In The Hague, on 22 aug 2018

Signature Entreprise Opf Bergheims Signature stagiaire _____
 Company stamp _____ Trainee's signature _____

*Merci pour votre coopération
 We thank you very much for your cooperation*