

BORDEAUX INP

PROJECT REPORT

FINAL YEAR PROJECT

Biomimetic control of a prosthesis from egocentric videos to reach objects

Author:

Marin GUTIERREZ

Cyril HATCHI

Antoine PITAUD

Mylene TAHAR

Supervisor:

Pr. Jenny BENOIS-PINEAU

January 2016



Contents

1	Introduction	4
2	General project presentation	4
2.1	Proposed solution	4
2.2	Project actors: LaBRI / INCIA	5
2.3	Robotic arm Poppy	5
2.4	Tobii glasses	6
2.5	GoPro cameras	6
3	State of the art and available knowledge on the visual search processing times	7
3.1	Free exploration	7
3.2	Motor response	8
4	Experimental protocol	9
4.1	Interest and goal	9
4.2	The protocol	9
4.2.1	Detailed protocol	10
4.2.2	Correction of the bias	10
4.3	Experiments and results	11
5	Data merging: a complex challenge	12
5.1	SDK Livestream Tobii	12
5.2	Live stream of a Gopro	12
5.2.1	Stream with two GoPro	13
5.3	Program to pilot the GoPros	13
5.4	Program launching the Gopro and Tobii live stream simultaneously	13
6	Saliency map over the 3D scene	14
6.1	Saliency map computation using Tobii	14
6.1.1	Existing program	14
6.1.2	Real-time processing	14
6.2	Gopro 3D scene retrieval	15
6.2.1	Stereo system calibration	15
6.2.2	Disparity map calculation	16
6.2.3	Associating the disparity map with the saliency map	17
7	Conclusion	18

List of Figures

1	The entire device on the patient	4
2	Robotic arm Poppy	5
3	Tobii glasses	6
4	Gopro 3D Dual Hero system	7
5	saccades direction	8
6	Proposed model	9
7	Tobii SDK architecture	15
8	Calibration grid	15
9	Epipolar geometry to calculate the 3D depth	16
10	Disparity map using block matching technique	17

Acknowledgments

We would like to thank **Pr. Jenny Benois-Pineau** who guided us through all this project and for her time.

We also would like to thank **M. Pierre-Marie Plans** who helped us at the begging of the project.

We also thank **M. Florent Paclet** and **M. Aymar de Rugy**, from INCIA, who allowed us their time and shared their experience of the project.

We thank as well **Dr. Manuel Boissenin** who assisted us and helped us working on the calibration theory.

Finally, we thank all the volunteers that came over to make the cognition video experiments and all the people that assisted us in any way with interesting discussions.

1 Introduction

Robotic prosthetics are today able to improve considerably the lives of disable persons. Researchers are still trying to make more easily usable devices, and one of the main remaining challenge is to make these prosthetics cheaper. In fact, a robotic arm can cost today up to 50000€, which is not affordable for most people. This project is aiming to improve the quality of a robotic prosthesis arm by using video processing coupled with neuronal signals sent from the brain of the patient.

There are here two main challenges that we have dealt with during this project. The first one was to convert the intention of the patient into numerical assumptions. The second one was to tackle the equipment issue for compatibility and synchronization of the different devices.

2 General project presentation

2.1 Proposed solution



Figure 1: The entire device on the patient

The solution that has been chosen is to use a prosthetic robotic arm Poppy for the mechanical part, driven by neuronal signals that are helped by data from video analysis. Fig. 1 shows the different devices that the patient is going to wear. In yellow is shown the location of the robotic arm, which ideally should not be amputated over the elbow. Then a GoPro stereo system is placed on the patient shoulder, on the side of the disable member, which is used for a 3D reconstruction of the scene. Finally, glasses are used on the patient to record his eyes fixations and to create a saliency map.

2.2 Project actors: LaBRI / INCIA



To realize this project, two main actors are involved: the LaBRI laboratory and INCIA (Institut de Neurosciences Cognitives et Intégratives d'Aquitaine). The LaBRI team that is involved is composed of students, research engineers and professors specialized in computer science and video analysis. The INCIA team is more specialized in the neuroscience part of the project.

2.3 Robotic arm Poppy

The Poppy project is an open source community that develops softwares, mechanics and electronics in order to make easily programmable robots. It is based on 3D printing thus the production of such robotic prothesis would be very cheap, compared to other possible materials.



Figure 2: Robotic arm Poppy

2.4 Tobii glasses

Tobii glasses are wearable electronic glasses that have multiple purposes. First, it has a frontal camera that records the egocentric scene that is in front of the patient that will be wearing it. It also has eye-fixation captors that records the eye saccades and microsaccades by saving the gaze fixations. Combining these two data, it is then possible to create saliency maps that represents the scene and the region of interest where the patient is looking at. It is possible to get these data either after a recording, directly via the SD card, but Tobii glasses also has a WiFi IPv4 system and an Ethernet IPv6 connection available to get the videos and data in live.



Figure 3: Tobii glasses

2.5 GoPro cameras

Gopro cameras are very small wearable camera in high definition, that were first created for making sports video, but they are now very popular and used for many different usages. They have an adhoc Wifi that can be used using a Gopro app (on cellphone or computer) to save video, get the video stream live, and also pilot the camera from a short distance. For this project, two Gopro Hero 3+ cameras are assembled using a GoPro 3D DUAL Hero system, as shown in fig. 4. This system makes it possible to use simultaneously the two cameras and to create a 3D view of the scene. It actually works based on a master and slave system where only one of the camera is receiving the commands, and the Dual Hero tool is sending the same request to the second Gopro. However we have encountered several technical issues using the Gopro and the 3D system.



Figure 4: Gopro 3D Dual Hero system

3 State of the art and available knowledge on the visual search processing times

In order to be able to determine whether the subject is focused on an object on the scene or not, a basic comprehension of the mechanisms which are involved in visual search are needed.

3.1 Free exploration

First of all, it is essential to understand that the processes involved in visual perception in humans are broad and various. This is why it is impossible to synthesize all its mechanisms with a single model.

Thus, the neural networks activated in response of the visual stimulus won't be the same according to the properties of the scene the subject is looking at: natural or not, known objects (de Mattos Pimenta and Tiedemann, 1990), partial or incomplete information (Farah, 1984), existing links between the object and the context (Nettleton et Bradshaw, 1983), and others.

Some of these mechanisms are top-down designed while others are bottom-up, some are serial and others are involving simultaneously several brain areas . . . It is easy to understand why the visual search processing times are so much dependent of the nature of the scene. However, some of these times have been found for key tasks in visual search. We have used the statements of Hélène Devillez's [9] and Denis Fize's work [10] as a basis for our model. During visual exploration, the eye-tracking methods make it possible to identify the two types of eye-movement: **saccades** and **fixations**. This process is serial and is repeating in order for the brain to build a consistent mental representation of the scene. Fixations have an average duration of **200-250** milliseconds but can vary from **100ms to 500ms** for the same subject (Rayner, 1978). The information needed by the brain to understand a scene is selected during these fixations times. Saccades are much shorter than fixations, but are not instantaneous either: it still needs **50ms to 100ms** in order to change direction (disengagement, saccade, and reengagement). Moreover, there is an average **latency** of **230ms to 280ms** for the first saccade.

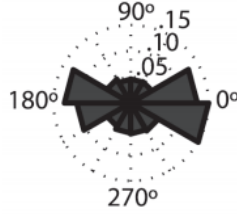


Figure 2 - 6 : Distribution des directions de saccades lors de l'exploration libre de scènes naturelles pendant 5 secondes (extrait de Ossandón et al., 2010).

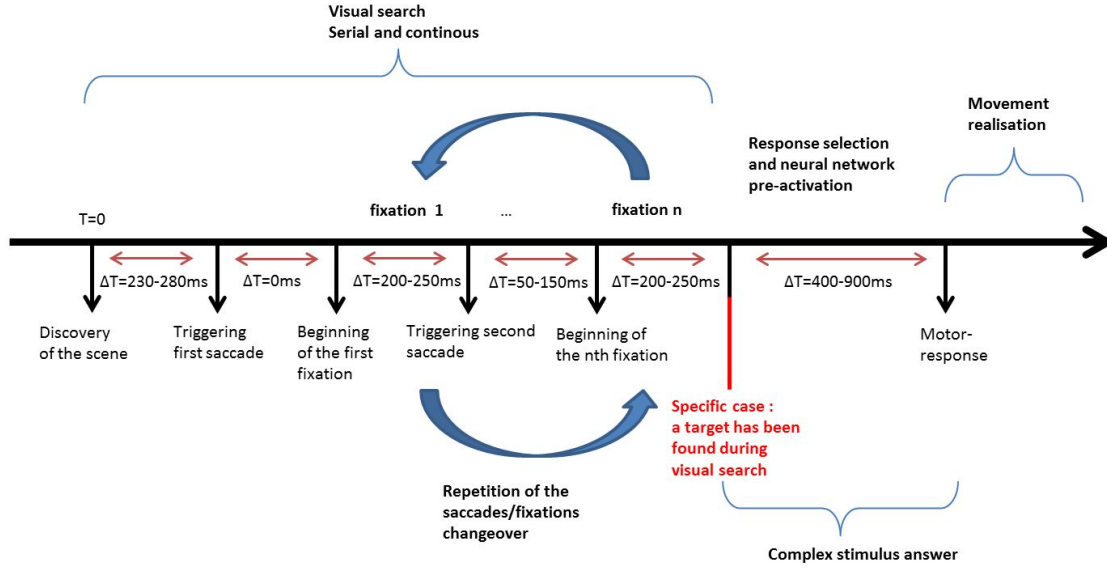
Figure 5: saccades direction

The detailed way the visual search is conducted needs others mechanisms to be explained. For instance, on figure 2, we can see average distribution of saccades during free exploration, favoring some directions and not others.

3.2 Motor response

While the visual search is processing, a particular object from the scene may draw attention and becomes a target we want to grasp. In order to be able to use the data from studies about reaction time, we will consider that this decision to grab and object is our starting stimulus. This allows us to transfer this problem to motor-response and movement preparation fields. Several works have been conducted to measure a reaction time, which is the duration between the activation of an expected stimulus and the answer of the subject. Two types of reaction times exist in the literature: simple (choice between two options) and complex, which implies a choice from the subject and a treatment before he can execute the chosen motor-response. Selected data will need to be compared to the one we will find in our study, which concerns the preparation of a complex movement. They will probably be longer, and it will be important to take care about the factors affecting these times: stress, tiredness or extra-individual specificity. Times found in literature are the following: 200ms to 250ms for a simple reaction time, and 400ms to 900ms for a complex reaction time, according to the situation and to the movement nature.

The following model can finally be proposed:



Temporal model for vision exploration

Marin Gutierrez
Cyril Hatchi
ENSC

Figure 6: Proposed model

4 Experimental protocol

4.1 Interest and goal

Once this theoretical model concerning visual search was established, we needed to validate it. Indeed, as explained in part 3, most of the studies measuring reaction time (RT) have a real strict framework, in order to be able to tell precisely if a measure concerns simple or complex RT, and especially so the beginning of the stimulus can be easily detected. That is why most of these studies uses a screen and a button : this way, they can deliver the stimulus at a precise time and they can know the exact moment when the subject reacted. But due to the context of the project, the goal of our experience was to test the visual search processing time in a real scene and not in a virtual one : that is precisely these kind of results which are hard to find in the literature. To summarize, we wanted to use the Tobii Glasses 2 in a controlled scene in order to determine the processing time of a subject for visual exploration and detection of an item.

4.2 The protocol

We proposed a protocol that we are going to describe here. Quite a few constraints needed to be taken into account for this experience: we had to use the Tobii Glasses 2 and the

Dual GoPro system. We also had a series of predetermined colored items which were going to compose our scene.

4.2.1 Detailed protocol

After the subject signed the consent of the experience, the first thing to do is to calibrate the eye-tracker and to position the DualGoPro properly.

Then, the experimenter gives to the subject the instructions : he will have to observe each of the proposed scenes and to grab the designated item with his or her right-hand as fast as possible.

The subject is blindfolded while the scene is set up.

The two scenes take place in the same set : an empty office facing a white and empty wall, which doesn't contain any elements which could potentially disturb the visual search. The items are placed at a convenient and predetermined distance for the subject : there is a total of 9 possible spots for 5 items at most. These spots are distributed on two parallel lines, one close to the subject and one a bit further.

4.2.2 Correction of the bias

We corrected some of our bias with our first subject. We had not consider that the sound of the items during the construction of the scene could help the subject in constructing a first representation of the scene, even if they were blindfolded. We tried to solve this problem with headphones, but they were not compatible with the Tobii Glasses and they affected its calibration. The solution we chose was finally to make some "false" noises in order to mislead the subject concerning the real position of the item. On the same level, the subject was influenced by the position of the experimenter, who is now in front of him when placing the items.

We first used five different items with five different possible position for them in the scene. To prevent the subject to deduct the position of the last item during the first step of the experience, we added another line where the items could be placed. However, we didn't used these recordings for our analysis.

We also realized a bit too late that the field of view of the Tobii Glasses' camera was too high : we had to asked to our next subject to voluntarily keep their head a bit lower than usual in order to have a view on the scene.

Finally, one bias was due to the color of the items. For example, the cylinder was the only blue piece of the set, so if we used it, as soon as we would have told "cylinder" to the subject, he would have searched the blue item.

As the color distinction is much faster than the shape distinction, it would have lead to errors in our analysis. Indeed, you can determine easily a color with your rods (without using your fovea), meaning you don't even have to look in the item's direction.

To solve this problem, we decided to work with only pairs of items, both having the same color but not the same shape. This way, when we ask to the subject to find an item, he can't rely on its color to find it quicker.

Objects kept for the scene are the following :

- tent and hexagon (yellow)
- square pyramid and cone (green)
- tetrahedron and parallelepiped (red)

4.3 Experiments and results

The establishment of the experiment took a great amount of time, considering we weren't aware of the best way to use the available equipment and that we encountered a lots of problems during the analysis. First, we had to conduct a pretest in order to eliminate the bias we talked about, but also to be sure that the data would be analyzable. For example, we had no feedback from the Dual GoPro system : the jacket on witch we strapped it is not really tight, and the view we are able to get was not good at all at first. We solved this problem by downloading the GoPro application, but the cameras' view still depends on the movement of the subject. The calibration of the Tobii Glasses are not as easy as what is announced. Despite all our efforts, one of our subject couldn't attempt to the experience, because we weren't able to calibrate correctly the eye tracker. We also have the feeling that this device is highly sensitive when it slips, even slightly, from the nose of the subject. For example, when we tried to use earphones in order to avoid sound spatialization, the simple contact with ears seemed to disturb the device.

The second main problem we met was to figure out how to extract and to analyze all the data we gathered from the Tobii Glasses. Several tools were available : the official software from Tobii (Tobii Pro Glasses Analyzer), their Software Development Kit (SDK) and finally an in-house developed SDK allowing to superimpose eye-tracking data with the video stream. Overall, the problem was quite the same for each of these tools : they are not precise enough for us to determine the delay between two fixations. The purpose of the Tobii Pro Glasses Analyzer is to show where the subject is mainly looking at, and does not permit to slow the video stream like we expected. Though, it allows to extract the raw data from the Tobii in a .TSV format, which can be opened with a simple spreadsheet software. We used these data in order to try to measure some interesting times, even if they couldn't be displayed with the video stream.

Note: Except from the raw video stream, there is no direct way to use the files contained in Tobii's SD card. A software has to be used in order to be able to exploit these data. The free licence for the Tobii Pro Glasses Analyzer lasts one month.

We are now going to introduce what we were able to determine from the data we gathered. As explained above, we used the .TSV file which gives, each 20ms, every information needed (gaze point, gaze direction, pupil position, and if the eye was found or not). The first thing we spotted was that the scene's complexity does has an effect on the visual search and the grip of the object. When the scene was composed by the five simple items, the time between eye opening and item prehension was on average 1.22 seconds longer than when the scene only had one item in it. We can also describe two kinds of simple patterns we spotted in the way visual search is performed.

First, the more used one is the one our model describes : one to three fixations are visible

along the scene before the eyes can spot the target. The Tobii Glasses are not precise enough, but there are probably several saccades on each item. Once the target is acquired ("a" fixation on the item), it seems that the following fixations are more centered toward the target, but not necessarily on the target itself.

On a less regular basis, we noted that sometimes, the subject doesn't seem to have a single fixation on the target. It may be caused by the imprecision of our Tobii, but we doubt it : there is no missing fixations and the movement is started before we can see a single fixation on the target. This observation throws into question the proposed model and the way we could detect the intention of movement.

5 Data merging: a complex challenge

A big part of the project is to use several sensors to be able to compute and send enough information to the Poppy arm. In fact we had to merge the GoPro system and the Tobii system in one operating system. This system is built to compute all the 3D scene and to recognize the interesting object in there.

To do this an SDK have been already developed like the SDK live stream for Tobii and a software which is able to generate the 3D scene. In fact we have to developed a tool to manage the GoPros and merge all the different software in one.

5.1 SDK Livestream Tobii

When we started the project, there was an engineer who worked few mounths on this project, M. Plans. The SDK for Tobii stream was developed by M. Plans, this SDK is able to display the video stream from the Tobii with the WIFI technology. Tobii uses the IPv4 protocol and the SDK works well. This SDK was developed before our arrival in the project team, and this SDK is written in C++ and uses lot of different packages. FFMPEG and OpenCV are used and we are able to stream in real time the movie camera on the Tobii Glasses and the information about the eye-tracking which are send by the Tobii. These informations are critical, the stream and the eyes-tracking information are used to recognize the object of interest in the scene. Actually the software is able to send the saliency map in the 2D scene but not yet in 3D, this is why we need to compute the 3D scene in pseudo-real-time, with that we will be able to compute the 3D-coordinate of the interesting object.

5.2 Live stream of a Gopro

Gopro is a small camera which can film with high resolution, it can diffuse the movie with the WIFI or save it on its own SD-card. A software developed by GoPro already exist, with it it's possible to watch the stream on a cellphone but not in a computer so we needed to write an application to watch the stream on the computer and to save every frame of each camera to compute the 3D scene.

Gopro has its own Wifi server and we can connect ourselves on it. Some functions are easy to find and to manage with Http requests. In fact, when the computer is connected with the GoPro, it is possible to change the status, start and stop the video stream, take a picture etc. For example, this sentence can change the status of the GoPro between movie

status or photography status, in our software the requests are send with the curl tools:

`http://10.5.5.9/bacpac/PW?t=MyPasswordl&p=%02`

This sentence shows that the IP adress of the GoPro is always 10.5.5.9 because each GoPro has its own server and the IP address never change.

To display the stream and save the frame we can't use http requests but it's easy to do with OpenCV and the object *VideoCapture* which is initialized with the url of the GoPro stream :

`http://10.5.5.9:8080/live/amba.m3u8.`

Opencv sees a stream as being a line of frame, with this way it's easy to save each frame and this is what we want. With this method we observed a latency between what the GoPro film and what we see on the computer, this latency is for four seconds.

In our project there are two GoPro which are combined with the 3D-DUAL HERO GoPro system wich use a slave/master protocol.

5.2.1 Stream with two GoPro

As mentioned beforehand, we use the 3D-DUAL HERO system and it is not compatible with our needs. Actually, with the Gopro 3D stereo system, one GoPro is slave and the other master, to keep the synchronization between these two we must be connected in WIFI with the master camera. The http requests are sent to the master Gopro which send them to the slave. In fact just one GoPro is connected (the master one) and we can display its stream but we can't display and save the stream of the slave GoPro which one is not connected. We contacted the GoPro support but, according to them it's not possible to display and save each frame from the stream without flashing the GoPro operating system. As a result, we decided to only work with one of the Gopro camera and the Tobii glasses, this problem will be tackled in further work. For instance, a dissociation of the two Gopro have been thought: the two Gopro would be piloted individually, from two different interfaces of the computer.

5.3 Program to pilot the GoPros

A program was written in order to pilot the Gopros directly from the computer. This was done by writting a GoProManager C++ class handling many functionalities such as: change mode, start capture, stop capture, start and save stream. This programs runs in a Ubuntu 14.04 LTS terminal and the commands are given by a user input directly into the terminal.

CURL library has been used, as GoPro cameras are piloted by HTTP commands received on its IP address.

This program works with one or two gopros, if they are coupled using the Gopro 3D stereo system. However, only the stream of the master camera can be saved on the computer.

5.4 Program launching the Gopro and Tobii live stream simultaneously

We wrote a program that launched the GoPro stream and the Tobii stream, two threads are used to do this. As it is said before we cannot recover the stream of two GoPros but just one. This main program used four parameters, two for the Tobii SDK and two for the GoPro manager:

- Tobii IP address
- Tobii Buffer Size
- GoPro IP address
- GoPro network password

In fact this program launches the GoPro Manager and the Tobii SDK, it has to be connected on the both networks, Tobii network and GoPro network, that requires two WIFI dongles. For this project and for the video processing, data from Tobii and GoPro must be synchronized and our main program doesn't do that and we have no solution for the synchronisation yet.

6 Saliency map over the 3D scene

6.1 Saliency map computation using Tobii

Tobii glasses has a system recording a eye-tracked movement. This actually means that the eyes positions are saved, and that it is possible to approximately know where the patient was looking at a certain time. This recent "eye-tracking" techniques are more and more used to modelize the position where a person is looking at and is following his gaze. As human vision is attracted by narrow regions in the scene (the saliency), we intend to use Tobii glasses in order to create a saliency map of the scene the patient is looking at using his gaze fixations, both these data are saved by Tobii.

The subjective saliency maps are computed using the D.S. Wooding method. This method construct a Wooding map that is based on the gaze fixations of the patient's eyes. Then, the saliency map is calculated based on the Wooding map.

6.1.1 Existing program

In previous work, an offline program has been created that constructs saliency maps after processing an excel file that is saved by the Tobii glasses. This program is using the gaze fixation data and the video from Tobii glasses. It also has an interpolation step to retrieve an entire sequence of gaze fixations, as many data of eye tracking are lost, because of closed eyes for example.

6.1.2 Real-time processing

The real-time processing is implemented using the Tobii SDK previously developed at LaBRI. It allows real-time retrieval of the gaze fixation data from Tobii glasses, as well as the video stream from the frontal camera.

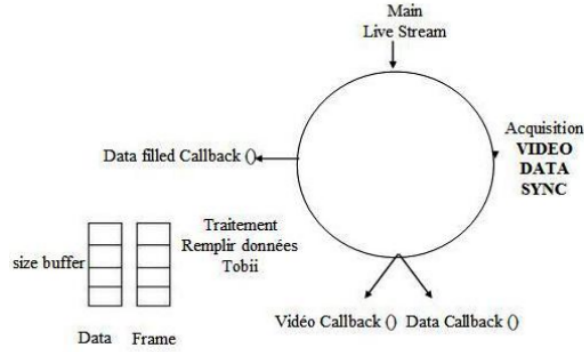


Figure 7: Tobii SDK architecture

As shown in fig. 7, the Tobii SDK makes use of callbacks for video and data. A synchronization step was also necessary as the data from the frames and the data from the gaze fixations are not received simultaneously. This is why the data received from the live stream are recorded on buffers, which size is to be defined by the user.

In all our tests, we generally used a buffer of size 10, which gave better results.

6.2 Gopro 3D scene retrieval

6.2.1 Stereo system calibration

The calibration step is a crucial part of the algorithm. In fact, the calibration not only corrects the image distortions, it also gives the necessary matrices to retrieve the 3D scene: the fundamental matrix of the optical system and the reprojection matrix, in order to get the real depth of a point in the scene. This step is done only once.

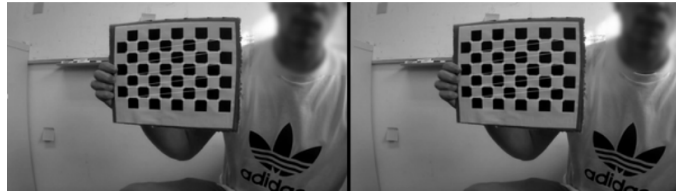


Figure 8: Calibration grid

Fig. 8 shows a calibration being done using a calibration grid with the Gopro stereo system.

The distortion is due to the optical system of the cameras, especially as we are using Gopro cameras that has a very short focal field. This distortion is to be corrected using the following method.

Two distinct types of distortion needs to be considered:

- The radial distortion which lead to a "fish-eye" effect, it means that lines on the border of the image are round-shaped.

This effect is corrected in x and y directions using the following formulas :

$$x_{corrected}^{radial} = x(1 + k_1r^2 + k_2r^4 + k_3r^6) \quad (1)$$

$$y_{corrected}^{radial} = y(1 + k_1r^2 + k_2r^4 + k_3r^6) \quad (2)$$

- The tangential distortion occurs because the image taking lenses are not perfectly parallel to the imaging plane. It is corrected correctly using the formulas :

$$x_{corrected}^{tang} = x + [2p_1xy + p_2(r^2 + 2x^2)] \quad (3)$$

$$y_{corrected}^{tang} = y + [p_1(r^2 + 2y^2) + 2p_2xy] \quad (4)$$

In the openCV method, these radial and tangential parameters k_i and p_i are presented as a row matrix:

$$Distortion_{coef} = (k_1 \ k_2 \ p_1 \ p_2 \ k_3) \quad (5)$$

After that, the following formula is used for the unit conversion :

$$\begin{pmatrix} x \\ y \\ w \end{pmatrix} = \begin{pmatrix} f_x & 0c_x & \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \quad (6)$$

Where f_x and f_y are the focal length of the system and c_x and c_y the optical centers, in pixels in the camera matrix. The calibration is aiming to retrieve those matrices.

The distortion correction is done by searching for a grid pattern in the image, as it is shown in fig. 8. This is done by memorizing a few patterns, before sending them to the calibration method with other parameters.

The calibration is first done separately on each of the two gopro cameras, before a global calibration of the stereo system is processed, which lead to a disparity-to-depth matrix, used in the next step.

6.2.2 Disparity map calculation

The disparity from a stereoscopic system represents the shifting of a pixel between the two views of the scene. We evaluate the distance in pixels difference between a point x in the left image and its corresponding point x' in the right image. The goal here is to retrieve the third dimension that was lost during the image acquisition: the depth of the scene.

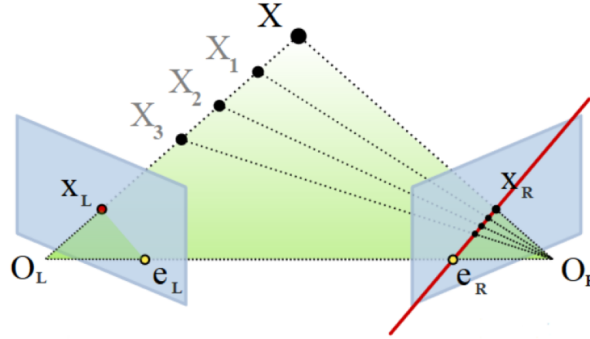


Figure 9: Epipolar geometry to calculate the 3D depth

As shown in fig. 9, the goal of the disparity map calculation is to find the exact position of the point X, at the intersection of the lines from images left and right. This representation is called epipolar geometry and can only be applied after the calibration is done, otherwise the lines would be incorrect because they would be curved.

Two possible algorithm are then possible to get the disparity map: Block matching algorithm or Belief propagation algorithm.

The first one, block matching, consists in searching for image blocks of 15×15 for instance in the second image (left) that correspond to the same view in the other image (right). This algorithm has the advantage of being very fast, but presents a few drawbacks, as it is shown in fig. 10.



Figure 10: Disparity map using block matching technique

The disparity on the ball is incorrect, this is due to the smoothness of its surface. This issue is tackled by covering the objects with stripes in experiments, but would still be to improve in further work, as a patient may encounter smooth objects.

6.2.3 Associating the disparity map with the saliency map

Finally, for the project purpose, it is necessary to consider that all these methods and techniques have to be working in real-time, as it has to be done to pilot the prosthesis instantaneously. Thus it is compulsory to think of fast algorithms that require only a small memory.

Then, the saliency map calculated from Tobii glasses and disparity map calculated from the Gopro 3D stereo system have to be assembled. To do so, a first step consists in retrieving spatially the shifting of the scene, which is different from the Gopro point of view (on the shoulder) and from the Tobii glasses point of view (on the head of the patient).

To make this correspondence possible, techniques as block matching or image registration can be considered.

7 Conclusion

It is essential to underline the deeply complex nature of this project. The difficulties encountered are both technical and physiological: the technology seems not completely mature yet, and the knowledge about visual search not precise enough. However, that is probably how research is progressing, so our efforts were not in vain.

After our work, we can suggest that in order to make a proof of concept, the detection of intention of movement should be triggered by the subject himself. Indeed, the detection of intention through eye-movement, if possible, can't be treated with the actual available device (Tobii Glasses). On the other hand, this device is adapted to point out, directly with the gaze, an item from the scene : it would be a much easier way to give simple instructions to the subject rather than to try to analyze his visual search, which can be deeply personal and varied from one person to another.

We also think it would be a good thing to contextualise the final product in order to get a closer idea on what the future users would want, in which context and with which conditions.

In way of conclusion, we can say we made a little step in a very complex environment, and that we explored quite deeply the proposed concept. We raised some issues about it, allowing the future researchers to focus on them.

References

- [1] Internship report of Ms. Mariem Cheikh. Mémoire "Analyse du flux vidéo pour l'amélioration des tâches de saisies par Neuroprothèses".
- [2] Internship report of M. Remi Delmas: "Acquisition et reconstruction de la carte de profondeur de la vidéo égocentrée pour l'asservissement d'une prothèse".
- [3] OpenCV documentation:

http://docs.opencv.org/2.4/doc/tutorials/calib3d/camera_calibration
- [4] Felzenszwalb, P. F., and Huttenlocher, D. P. (2006). Efficient belief propagation for early vision. *International journal of computer vision*, 70(1), 41-54.
- [5] da Costa, C. M. C. (2011). Multiview object recognition using Bag of Words approach. *Computer Vision (WACV)*.
- [6] Montesano, L., and Lopes, M. (2012). Active learning of visual descriptors for grasping using non-parametric smoothed beta distributions. *Robotics and Autonomous Systems*, 60(3), 452-462.
- [7] Markovic, M., Dosen, S., Cipriani, C., Popovic, D., and Farina, D. (2014). Stereo-vision and augmented reality for closed-loop control of grasping in hand prostheses. *Journal of neural engineering*, 11(4), 046001.
- [8] Degani, A. M., Danna-Dos-Santos, A., Robert, T., and Latash, M. L. (2010). Kinematic synergies during saccades involving whole-body rotation: a study based on the uncontrolled manifold hypothesis. *Human movement science*, 29(2), 243-258.
- [9] DE GRAEF, M. P., ALEXANDRE, M. F., DELORME, M. A., COURBOULAY, M. V., GUERIN-DUGUE, M. A., and GUYADER, M. N. Etude des processus attentionnels mis en jeu lors de l'exploration de scènes naturelles: enregistrement conjoint des mouvements oculaires et de l'activité EEG.
- [10] Degani, A. M., Danna-Dos-Santos, A., Robert, T., and Latash, M. L. (2010). Kinematic synergies during saccades involving whole-body rotation: a study based on the uncontrolled manifold hypothesis. *Human movement science*, 29(2), 243-258.