Faculdade de Engenharia da Universidade do Porto

Capture and Analysis of the Trajectories of Anatomical Points on the Face to Support and Evaluate Reconstructive Plastic Surgery

Ana Filipa Domingues Gerós

MASTER THESIS

Integrated Master in Bioengineering

Supervisor: Paulo de Castro Aguiar
Co-Supervisor: Jaime Cardoso

July 2015

Resumo

A avaliação da eficácia de terapias médicas e cirúrgicas, implementadas na área da cirurgia plástica reconstrutiva, tem vindo a aumentar de importância no que diz respeito à comparação e escolha do tipo de procedimento médico mais apropriado para lidar com casos de desfiguramento facial. No entanto, os métodos tradicionalmente utilizados para avaliar os resultados das intervenções cirúrgicas e, portanto, fornecer uma caracterização detalhada das expressões faciais, são quase exclusivamente qualitativos e falham na objetividade e na presença de informação dinâmica. Para além disso, os escassos métodos desenvolvidos com o propósito de quantificar objetivamente a eficácia destas intervenções têm pouca utilidade na área clínica por imporem fortes constrangimentos em termos de configuração de instalação, custo, complexidade de utilização e pouca adequação das métricas produzidas. Estas limitações impõem uma necessidade urgente de criação de um novo sistema para quantificação estática e dinâmica da face, rico em medidas clinicamente relevantes e que permita uma fácil interpretação, e consequente adopção, pela comunidade médica.

Desta forma, este trabalho teve como objetivo desenvolver um método capaz de, quantitativa e objetivamente, avaliar os movimentos faciais, usando um conjunto de medidas morfológicas, dinâmicas e estáticas. Este novo sistema, aqui designado por *facegram,* pretende ser uma ferramenta de baixo custo, simples e adequada para integração em ambiente médico. Este trabalho teve a importante característica de ter sido desenvolvido em estreita colaboração com o cirurgião plástico Ricardo Horta, do Hospital São João. Para atingir os objetivos referidos foram utilizadas câmaras RGB-D com tecnologia *Kinect* para aquisição das imagens. A utilização deste tipo de sensores tem muito potencial em métodos de análise de movimentos faciais uma vez que a caracterização destes movimentos complexos apenas é possível utilizando técnicas de medição a 3 dimensões. A metodologia proposta inclui um conjunto de algoritmos desenvolvidos para seguir a trajetória de pontos anatómicos faciais, através de um algoritmo do tipo *Block Matching,* algoritmos de pré-processamento RGB-D para melhorar a qualidade deste seguimento, modelo cinemático interno de forma a diminuir o tempo computacional e um algoritmo para correção do movimento da cabeça. Os resultados de desempenho deste conjunto de técnicas foram promissores, com elevados coeficientes de determinação e baixos erros médios, o que permitiu concluir que os métodos escolhidos são adequados para atingir os objetivos propostos. Estes métodos foram validados com registos RGG-D de 2 pacientes e 2 sujeitos de controlo. Os algoritmos foram encapsulados em dois programas com interface gráfico e de fácil utilização em ambiente clínico, para suportar, respetivamente, a aquisição e análise de registos.

Após implementação deste conjunto de técnicas para o seguimento dos pontos ao longo do movimento facial, os resultados foram organizados e as medidas mais adequadas foram

apresentadas no *facegram*, tendo em conta a sua utilidade clínica e fácil interpretação médica. Embora de *design* preliminar, o conteúdo do *facegram* foi aprovado por especialistas. A partir dos resultados obtidos para os pacientes e controlos foi possível concluir que o sistema proposto aparenta ser uma solução adequada para a caracterização objetiva e quantitativa dos movimentos faciais, com potencial para se tornar uma ferramenta standard e universal na avaliação dos resultados de cirurgias plásticas reconstrutivas.

Abstract

The evaluation of the effectiveness of existing medical and surgical therapies, in the reconstructive plastic surgery field, has increasingly become an important asset of consideration in comparing and choosing the most suitable type of medical procedure to handle the facial disfigurement. Unfortunately, the traditional methods to assess the results of surgical interventions, and so the detailed characterization of facial expressions, are mostly qualitative and lack of objectivity as well as dynamic information. Along with this, the few existing methodologies tailored to objectively quantify the effectiveness of these interventions are not practical in the medical field, since they impose severe constraints in terms of installation configuration, cost, complexity and poor suitability of the produced metrics. These limitations enforce an urgent need for the creation of a new system to quantify facial movement and allow for easy interpretation by medical experts.

With this in mind, the present research aimed to develop a method capable of quantitatively and objectively assess complex facial movements, using a set of morphological, static and dynamic measurements. This new system, proposed to be named *facegram,* is intended to be inexpensive, simple and, so, suitable to be integrated in the medical field. This work had the important feature of being developed in collaboration with the plastic surgeon Ricardo Horta, from *Hospital São João* (Porto). For this purpose, the RGB-D cameras with *Kinect* technology were used to acquire images. The use of these sensors has much potential in face analysis-based systems since a complete characterization of facial movement can only be achieved through 3D measurement techniques. The proposed methodology includes a set of algorithms developed for facial anatomic points' tracking using an algorithm of *Block Matching* (BMA), RGB and depth pre-processing to improve tracking, simple and internal kinematic model to accelerate computations and an algorithm to correct head's movements. The obtained results regarding tracking's performance were promising with high coefficient of determination values and low mean errors values, concluding that the chosen tracking method is suitable for achieving the established objectives. These methods were validated using RGB-D data from 2 patients and 2 controls. The algorithms were grouped into 2 programmes with a Graphical User Interface (GUI) to be easy-to-use in clinical environment, in order to support, respectively, acquisition and analysis of data.

After implementation of the methods to track facial points, the outcomes found after points' tracking were organized and the most suitable measurements were shown in the *facegram,* taking into account the clinical utility and the facility for medical interpretation. Although with preliminary design, the structure of the *facegram* was approved by specialists.

Based on the results for both patients and controls, it was possible to conclude that the proposed system appears to be an appropriate solution for quantitative and objective characterization of facial movement and with great potential to become a standard and universal tool in the assessment of reconstructive plastic surgeries' outcomes.

Acknowledgments

This work would not be complete without the help of all people who, directly or indirectly, "crossed my way" and support me through all these months (or even years). Above all and for all of them I leave the most heartfelt thanks.

First and foremost, to Paulo, since without him nothing of this would be possible! Thanks for all the support and motivation during these months and for helping me "opening some doors and windows" in my future life. But mostly, thanks for all the moments of extreme stupidity and laughter because, without them, this project would not have been so funny (PS: I hope you do not forget that, for a moment, you were a psychiatric patient and that this project ruined my modelling career!). I'm also grateful to Professor Jaime for all the helpful ideas and advices, wherever needed. Moreover, I have to thank to Dr. Ricardo Horta, from Hospital São João, for the medical support and for being so enthusiastic during the whole project!

Apart from the people who directly contributed to this work, many others that, "just by sitting there", already made a difference and to whom I deeply thank. In particular, for those who enter with me and joined me during these five years; particularly: To MAPADIS '10, for all unforgettable moments - without them all these years would be a "valente seca!": for all the music, nights, laughs and tears we share (an "*avé*" to Praxis!). In fact, some of them deserve special attention. To *Tigresas*, the "secret" group, for being there whenever is needed (and because it is never too late to meet some of the best people in the world!). And, especially, to *BFFs*, for the wholehearted support, 24 hours a day, 365 days a year: to Joana-X for all the crazy stuff and discussions; to Clara, for being the fastest housemate that I've ever had; to Mary, for the "caipirinhas", "FeupCafes" and "works" in bars, without which I would be completely crazy; to Pi, for all the advices and "cusquices"; to Isa, for making my life more bouncy, and the last but not the least, to Joana, for being my soul mate (and more words for what? You even know what I'm thinking, all the time!) and to Lia, my eternal companion of classes, jobs, cries, laughs, jokes and crazy things (because you know I cry when you are not with me, so no need to further explanation). Thank you all for making me a better person!

There are still some people who made my life happier since they appear. Thanks Filipa and Tiago, who, since 1995, rejoice my days and bore me so much in a way that I really like! To VENTI TRÈ, for sharing with me the best times I've ever had. To Xena and Soluços, for being the best godchildren ever: for all *Portos* and *Fados* we share (and for those we do not share!) and for letting myself be a "madrinha-galinha". Finally, to Amanda, the best praxis-godmother I could have (the affairs council already misses you!).

A special thanks goes to Zé João, who taught me that life is not only books (also beer!) and that is possible to transform a nerd into a normal person. Also taught me the true answer to the question "how is it?" (which, according to the expert, is: "it's like the black one but white!"). Above all, for having enough patient to “put up with me” in times of stress, for making an effort to say cheesy things, for all the sacrifice to warm my feet and for believe that it is possible to like “M” or ∞.

My gratitude goes also to the best teacher that I’ve ever had, Professor Marques, for showing me that “*He who ceases to be better, ceases to be good*” (Oliver Cromwell). Besides that, thanks for making me think about my way and my job in this world, and helping me to find one of my great passions – teaching.

Finally, to my mother, for being the person who upsets me the most but who supports me, no matter what (because being a mother is being you); to the great philosopher and humorist of my life, my father, for the deep reflections and, above all, for believing in my choices and be an inspiration (mainly, riding his bike); to my brother, for being a huge pain in the ass (but a person I love; I think that, deep down, you know that); to Nés and Aunt Ameixa, for being the best borrowed-sister and -mother that I can have and for all the love and “honey” since ever. Finally, to the rest of the family that loves to surprise me on special days!

Ana Gerós

"You have been told also life is darkness,
And in your weariness you echo what was said by the weary.
And I say that life is indeed darkness save when there is urge,
And all urge is blind save when there is knowledge,
And all knowledge is vain save when there is work,
And all work is empty save when there is love."

The Prophet, by Khalil Gibran

Contents

List of Figures

List of Tables

Acronyms

2D	Two-Dimensional
3D	Three-Dimensional
AAM	Active Appearance Models
AAO-HNS	American Academy of Otolaryngology-Head and Neck Surgery
AFA	Automated Face Analysis
ASM	Active Shape Models
BMA	Block Matching Algorithm
CLM	Constrained Local Model
DAS59	Derriford Appearance Scale 59
DBN	Dynamic Bayesian Network
DE-CM	Multiple Differential Evolution-Markov Chain
DLP	Digital Light Processing
DMF	Deformable Model Fitting
EQ-5D	Euroqol
FACE	Facial Assessment by Computer Evaluation
GUI	Graphical User Interface
HB	House-Brackmann
HUI	Health Utilities Index
IR	Infrared
KLT	Kanade-Lucas-Tomasi
LC	Left commissure
LMP	Left midpoint
LoG	Laplacian of Gaussian
LP	Linear Predictor
MAD	Mean Absolute Difference
MBSRQ	Multidimensional Body-States Relations Questionnaire
MSE	Mean Squared Error
MSRA	Maximal Static Response Assay
NUI	Natural User Interface
OSCAR	Objective Scaling of Facial Nerve Function Based on Area Analysis
PCA	Principal Component Analysis
RC	Right commissure

RGB-D	Color and depth
RMP	Right midpoint
ROI	Region-of-interest
SDK	Software Development Kit
SIFT	Scale-Invariant Feature Transform
SMILE	Scaled Measurement of Improvement in Lip Excursion
SVD	Singular Value Decomposition
Tof	Time-of-Flight
VCA	Vascularized Composite Allograft

Chapter 1

Introduction

Reconstructive and aesthetic plastic surgery is a specialty that adapts surgical principles in order to achieve the specific needs of each patient, doing remodelling and redistribution of the tissues. According to the *American Board of Plastic Surgery*, it handles the excision, repair, replacement and reconstruction associated to shape and function defects of the integument and underlying musculoskeletal system. Situations approached by reconstructive and aesthetic plastic surgery include birth defects, developmental abnormalities, trauma and posttraumatic pathology (burns and its sequels). In the particular case of cranio-maxillofacial surgery, which operates in the head, neck, face, jaws and hard and soft tissues of the oral regions, one of the aims is to circumvent the effects of facial disfigurement, also caused by cancer, trauma, neurofibromatosis, burns, congenital malformations or facial paralysis [1]. A survey study started in 2007, supported by the *Changing Faces* credited charity organization and with the aim of better estimate the impact of having disfiguring facial conditions, has found that 415,500 people in the UK are expected to acquire facial disfigurements in the period of one year and, at the present, 542,000 people have a significant disfigurement, in which 100,000 are from facial paralysis, including strokes [2]. However, these numbers are lower regarding the quantity of plastic surgeries performed to offset these problems.

Facial paralysis, being the causal factor that exhibits the greatest incidence rate, presents a challenging reconstructive problem, in which the operative treatments remain in constant progress. The current techniques, in which ***the goal is to restore facial symmetry both at rest and in active expression as well as to recover the spontaneous and natural smile*** (called facial reanimation techniques), are constantly being refined [3-5]. Nerve and muscle transposition, reinnervation and static techniques are examples of therapies that are currently used to restore the normal function of the facial structures [4, 5].

Apart from the functional problems that these medical procedures might resolve, such as speech impairment, problems with eating and closing the eyes effectively, they will also improve significantly the quality-of-life of patients regarding social interaction and communication. Studies suggest that facial disfigurement, like facial paralysis, can have a significant emotional, economic and social impact on patients, being more likely to report symptoms of depression, anxiety and hostility, and leading to psychological and physical fatigue [6-9]. Besides that, the disfiguring condition is essential for efficient communication,

since the dynamic of the conversation is greatly affected by facial movements as well as facial expressions [10].

1.1 Motivation

Along with the panoply of different techniques in recovering the normal facial movements comes the difficulty of selecting the most appropriate procedure for each patient and, in parallel, to assess the effectiveness of the reanimation techniques. Also the lack of available tools that are universal and consensual makes it impossible to quantify and compare the results of different surgical techniques [5, 11].

To achieve this and to standardize the evaluation of several therapies, it is necessary to have a precise, objective and consensual measurement system of the static and dynamic facial features [5, 11, 12]. Such quantitative and standard assessment must include components such as acquisition of patient's data (set of standard facial expressions) in two-dimensional (2D) or three-dimensional (3D) space, and spatial-temporal data processing for the analysis of facial point's trajectories, in order to improve the medical analysis and interpretation.

Despite the extensive research in the tracking of anatomical landmarks and facial reconstruction fields, the current evaluations of the results of surgical interventions are exclusively done using qualitative or semi-quantitative methods [12-19]. Common procedures include grading techniques [13, 14, 18], questionnaires, with psychological assessments [12, 15, 16]. However, the evaluation using these methods is subjective and ambiguous, and do not provide information about dynamic facial movement [5, 11, 20]. In the particular case of the questionnaires, it also has the disadvantage of depending on the judgment of the physician and patient opinions [15, 17].

To overcome these problems, various studies are being undertaken in order to develop quantitative methods capable of helping the physician to evaluate a patient's progress before and after surgery, and compare the effectiveness of different operative and medical procedures with precise determinations of the facial expressions [11, 21-35]. Nevertheless, the lack of dynamic and/or 3D information provided by these methods, as well as the use of time-consuming techniques, or even the cost of the imaging equipment, makes them ineffective and impractical in the clinical area [20].

For these reasons, there is a strong need for a quantitative method which is simple to install and to use in a clinical environment, non-expensive and capable of providing rich information for the evaluation of the effectiveness of the surgical and physical techniques [20].

1.2 Objectives

One of the main purposes of this research is the development of a new method to objectively and quantitatively assess the movement's properties of selected facial anatomical landmarks. This methodology was structured taking into consideration the motivations mentioned before, as well as the requirements to be effortlessly applicable in the clinical field.

Since the currently available methods do not allow a simple and complete quantification, through medically useful measures, it is also intended to create a set of morphological, static and dynamic, measurements and combine them into a report, easy to interpret by a physician and allowing a complete characterization of facial movements.

As the extremely complex facial movements can only be adequately and objectively described using a spatiotemporal analysis, combining tridimensional space and time, RGB-D cameras were selected for data acquisition. Low cost *Kinect*-based cameras were used in this work but the algorithms/softwares for movement analysis have been made general to allow RGB-D streams acquisition with different cameras (with different resolutions and different frame rates).

In order to achieve the proposed global aims, specific goals were established:

- Define a protocol for RGB and depth images acquisition to be applied in a clinical environment.
- Develop and implement a methodology to track facial anatomical points over time, as well as pre-processing techniques and an internal kinematic model to, respectively, improve contrast and accelerate the computations.
- Construct two Graphical User Interfaces (GUI), intended to be used in data acquisition and tracking of facial points, providing a user-friendly environment to facilitate the analysis process.
- Evaluate the performance of the developed tracking method by comparing the obtained results using an adequate validation system.
- Develop and construct a *facegram*, i.e., novel strategy to display the core variables, chosen to be the ones that best characterize facial movement. This system is intended to help the physician to quantify these facial excursions and, ultimately, evaluate the effectiveness of surgical techniques.
- Apply the developed methods to two groups, one of patients with facial paralysis and the other of controls, and obtain validation from plastic surgery clinicians.

1.3 Contributions

The proposed work had the following main contributions:

- Development of an application for RGB and depth images acquisition in a clinical setting, using a *Kinect*-based camera (RGB-D sensor).
- Implementation of a set of methods for determining the position of specific facial anatomical points over time. These methods include a modified *Block Matching Algorithm* (using also depth information) for tracking, color and depth pre-processing, correction of head's movement algorithm and a simple internal kinematic model.
- Development of a second application for tracking the anatomical points in medical environment. After tracking, its GUI presents the trajectory of each facial point, constructed with the raw data, to preliminary analysis.
- Creation of a new strategy for data presentation chosen to quantify the facial movement, i.e., from the reconstructed trajectories of the selected anatomical

landmarks, a selection of which core variables should be displayed and how should they be presented was carried out.

- Generally speaking, a novel method to perform tracking of facial anatomical landmarks and to quantify their movement in a complete, clinically relevant, dynamic and easy-to-interpret way was created, taking advantage of RGB-D sensors.

1.4 Document Structure

The first chapter provides an overview on the subject to be studied as well as the main purposes, objectives and contributions of the present work.

Besides this introduction, the document contains four more chapters. Chapter 2 describes the face anatomy and key-features that have to be taken into account for the review on the state-of-the-art that is subsequently described, addressing the specific subject. In Chapter 3, the methodology used to achieve tracking of facial anatomical points as well as the discussion of the main obtained results are described. The collection of the best measures and graphical representations for quantitatively assess facial movement (*facegram*) is described in Chapter 4, together with the results obtained for a normal subject and both patients. Finally, the main conclusions of this research are presented in Chapter 5.

Chapter 2

Literature Review

In this Chapter, the facial anatomical and physiological concepts, complications and the medical procedures used to circumvent these problems will be presented. Additionally, the study and state-of-art of several methods that are, or could be, used to evaluate the effects of these procedures will also be analyzed. Finally, the intermediate stages commonly implemented and widely studied to address this purpose and, more properly, to quantitatively assess the outcomes of reconstructive plastic surgeries, will be considered.

A thoroughly revision of the literature has been performed and this chapter describes in detail many of the publications relevant for this work. A condensed perspective of the field can be obtained by reading only sections 2.1 to 2.3 and the Summary sub-sections of 2.4 to 2.7.

2.1 Anatomy and Physiology of the face

The human face is an extremely complex and specialized organ, responsible for transmitting sensory information to the cortex, received from the environment. In addition to biological functions, the face also has a social dimension, insofar as it accounts for expressing emotions, playing an essential role in the interpersonal communication. Interestingly, the face only became an "organ" on July 3rd, 2014, and this term applies solely in the medical transplantation world, as a result of changes in the *US Federal health* policy. Regarding the fact that the face transplantation was introduced into the jurisdiction of medical procedures, a new term was created, defining the face as a vascularized composite allograft (VCA) organ, i.e., a complete sections of tissue that are removed for transplantation [36]. Besides face transplantation, the other equally common VCA procedure is the hand transplantation.

The anatomic structure of the face, although complex, can be divided according to the tissue composition: skin, soft tissue (including fat, muscles and connective tissue) and hard-tissue components (bone, teeth and cartilage), which, all together, form the three-dimensional shape of a particular face [37, 38]. In this sense, the face appearance depends on the convexities and concavities of the underlying skeleton, to which the hard and soft tissues are attached [38, 39]. More particularly, the facial contour is mainly delineated by the

convex facial bones, namely the nasal bones, the supraorbital borders, the zygomatic bones, and the mandible [37] (see Figure 1). These structures, as well as the remained ones, that constitute the bone anatomy of the face are shown in Figure 2.

Although these structures are universal for all human beings, the differences in the referred tissues and the variations in the coloration and texture of the skin and eyes are responsible for the identification of each individual's face and his singularity so characteristic of the humans. Besides that, according to the race and ethnicity of each subject, different features are visible, resulting from anatomical changes, characteristic of these different groups [37, 38].

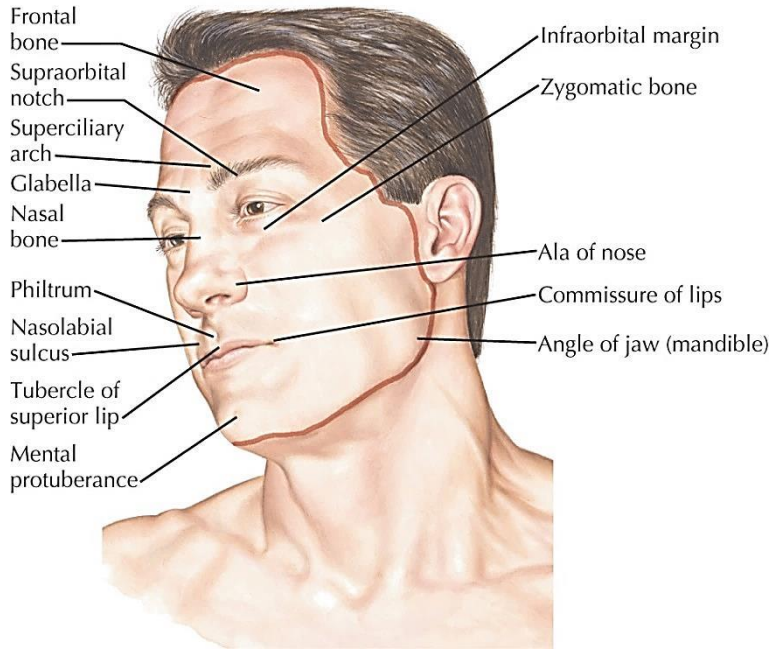


Figure 1. Topography of the face anatomy. Adapted from [40].

Regarding the fundamental functions of face [looking, eating, breathing], the presence of normal sensation originated from the process of receiving sensory information from the environment, is important for the discrimination of temperature, touch and pain, and plays an important role in monitoring and self-defence situations [41]. Besides that, human face is responsible for expressing emotion, conscious and unconsciously, providing and receiving important social signals, critical to the interaction with other individuals [38]. The face is, therefore, a critical element in human communication and social interactions. Even mild facial disfigurements can have a devastating impact in the patients' quality-of-life [6-8, 10].

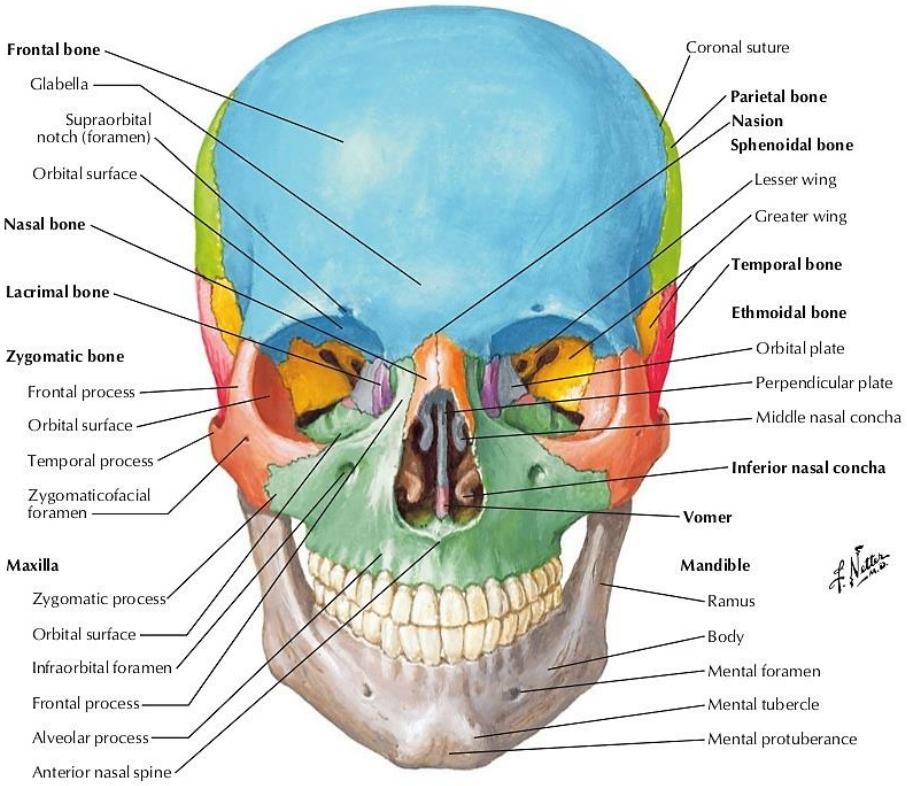


Figure 2. Bone anatomy of the face. Adapted from [40].

It is the movement of the face that produces these social signals, i.e., expressive movements, liable for providing information that improves speech perception. There are two types of facial motion, rigid and non-rigid movements. The former is related to movements that do not change the shape of facial features. Example of these is the movement of the head that usually signals the viewing direction. Non-rigid movements affect the facial features and are the main responsible for social signals, such as speech and representation of feeling and mood signals [38].

All movements necessary to perform the different functions of the face require the activation of muscles, or system of muscles, distributed over the face. The facial musculature can, then, be divided into muscles of facial expression and mastication muscles, and additional information about each muscle type can be found in the literature [37, 39, 42].

2.2 Facial disfigurement

Since faces are complex body structures, with different constituents that are interlinked, their functions or appearance can be changed in numerous ways. These modifications that reflect an harmed appearance with visible anatomical differences, and involve the physical and psychological health, are called facial disfigurements [43]. A wide variety of factors can cause facial disfigurement and range from the most frequent, such as cancer, trauma, facial paralysis or burns, to the rarest, like congenital malformations [38, 43]. In fact, the causes can be segmented into two categories. The acquired disfigurements, as the name suggests, include those resulting from trauma, abnormal developmental processes, surgical interventions or even diseases, which occur at a given moment of the patient's life. On the

other hand, congenital malformations relate to conditions that cause visible differences that are manifested at birth or remarkable over time.

Regardless of the abnormality source, all of them are responsible, somehow, for a distortion in the natural symmetry of the face. In fact, numerous studies have shown that the face attractiveness is closely related to facial symmetry, having a positive impact in social relationships [44-46].

In addition to problems associated with the aforementioned vital functions of the face, there is a negative social impact related to facial disfigurements and the patients that suffer from that are more likely to be socially constrained and isolated, what results in low self-esteem and psychological distress [6-8].

2.3 Reconstructive and aesthetic surgical procedures

In order to restore the symmetry and mitigate anomalies of the face, as well as re-establish the full functionality and normal appearance of the individual, techniques of reconstructive plastic surgery are used, and the efficacy and safety of these approaches are based on the extensive knowledge of facial anatomy [4, 39, 47].

Although being closely related specialties, the aesthetic surgery and plastic surgery are not defined in the same way, regarding the medical field. While the aesthetic or cosmetic surgery relies entirely on improving the appearance of the patient with respect to symmetry or proportion, plastic surgery is dedicated to reconstruction of body and face defects or anomalies, correcting the dysfunction caused by the affected area [48]. In the latter, in which the present study is entirely focused, there are in-numerous procedures to reconstruct facial deformities, namely, cancer reconstruction, cleft lip and palate repair, facial paralysis, facial trauma reconstruction, nose reconstruction, scar revision and even face transplants [49, 50].

From all the causal factors, facial paralysis, i.e., the loss of movement by some or all the muscles on one side of the face, is the one with highest incidence rate, affecting approximately 70 in 100,000 people, worldwide, and it was estimated the occurrence of 127,000 cases of permanent paralysis annually [4]. There are many different causes of facial paralysis, such as infection, trauma, toxins, and idiopathic, neoplastic, congenital and neurological causes, all of them resulting from damage of the facial nerve or to the area of the brain responsible for the movement of the facial muscles [4, 51]. A condition associated with facial paralysis, also requiring clinical intervention, is synkinesis. This one involves abnormal involuntary movements, which occurs during a voluntary movement, from a different muscle group [26].

The selection of most appropriate to surgically correct facial disfigurement treatment is done according to the affected area and the degree of abnormality, always with the aim of ensuring a return to functionality and initial appearance, which are truly important to the quality-of-life of patients, both at functional and social level.

2.4 Evaluation of surgical interventions

Despite the multitude of facial treatments that can be performed, surgeons are faced with difficulties in properly assessing their outcomes and, ultimately, to choose the most

appropriate treatment for the patient based on the comparison between the effectiveness of different techniques. The outcomes assessment are related to the evaluation of the results for surgical interventions, prioritizing the effectiveness of the outcomes based on patient's satisfaction and the costs of reaching the expected results [17]. The primary purpose of evaluating the treatments' outcomes is to have solid evidences, based on standardized measurements, on which to base medical decisions, more specifically, to allow the comparison of techniques or between results for an individual subject, identification of patients for whom certain surgical procedures are not beneficial and finally quantification of positive results [16, 17].

However, currently there is a lack of universal and consensual instruments that can be widely used for outcomes evaluation in the realm of facial plastic and reconstructive surgery [5, 11, 15, 17]. A situation that contributes to this lack of methods is the difficulty in organizing and collecting data efficiently, in outpatient clinic environment [17].

With all of this in mind, a precise, objective and consensual measurement system, capable of standardize the evaluation of several therapies, would be extremely useful in quantifying the treatments outcomes, probably becoming a universally accepted method for clinical practice.

Currently, and despite the progress in this field, the evaluation of the results from interventions in reconstructive plastic surgery of the face is performed mainly using qualitative and semi-quantitative analysis.

2.4.1 Qualitative and semi-quantitative methods

Qualitative methods can be further divided into two categories: grading systems and questionnaires.

The grading systems propose an evaluation of the facial features with respect to the movements of the muscles and assessment of symmetry, at rest and in motion. The most common and widely used is the *House-Brackmann* (HB) system, which classes lower motor neurone facial nerve palsy, assigning 1 of 6 set categories (Table I) to the patient, which presumably describes his facial motor function and secondary defects [11, 14, 18]. In fact, this system was renowned as the universal and standard tool for the evaluation of facial paralysis by the *Facial Nerve Disorders Committee of the American Academy of Otolaryngology-Head and Neck Surgery* (AAO-HNS) [52]. Although the scale contains sequential ordered grades, it is not uniformly distributed, since the distance between each grade is different. In addition, this system is not sufficiently sensitive to small changes in facial expressions, so that it is not possible to completely characterize the facial movement. Finally, since the scale's grades are arbitrary, the data cannot be mathematically manipulated to yield objective findings [14, 18].

Other methods have been developed as a way to address the limitations of the HB system. The *Sunnybrook Facial Grading System* developed by Ross et. al. [13] is based on the evaluation of resting symmetry, symmetry of voluntary movement and synkinesis. The latter two are based on movements produced by 5 standard facial expressions, in which a score of 1 to 5, or 0 to 3, is assigned, depending on the degree of movement and synkinesis, respectively, reported by the physician. The resting symmetry category is classified, with tabulated values between 0 and 2, taking into account the features of three distinct locations of the face (eyes, cheek, mouth) when compared to normal side. Finally, the patient's

classification is made according to the grades in each category [13, 18]. The *Sunnybrook Facial Grading System* is depicted in Figure 3.

Another commonly used method, and that it is also based on grading systems, is called *Sydney Facial Grading System*. Unlike the previous, this is based on the evaluation of the voluntary movement of 5 anatomically relevant branches belonging to the facial nerve and a grade from 0 to 3 is assigned to each one, depending on the amount of movement. In addition, it is also assessed the synkinesis of the overall face, on a scale from 0 to 3 [18] (Figure 4).

Table I. *House-Brackmann* facial nerve grading system. Adapted from [14, 18].

Grade	***Description***	***Characteristics***
I	Normal	Normal facial function in all areas
II	Mild dysfunction	Gross: slight weakness noticeable on close inspection; may have very slight synkinesis. At rest: normal symmetry and tone Motion: Forehead: moderate to good function Eye: complete closure with minimum effort Mouth: slight asymmetry
III	Moderate dysfunction	Gross: slight weakness noticeable on close inspection; may have very slight synkinesis. At rest: normal symmetry and tone Motion: Forehead: moderate to good function Eye: complete closure with minimum effort Mouth: slight asymmetry
IV	Moderate severe dysfunction	Gross: obvious weakness and / or disfiguring asymmetry At rest: normal symmetry and tone Motion: Forehead: none Eye: incomplete closure Mouth: asymmetric with maximal effort
V	Severe dysfunction	Gross: only barely perceptible motion At rest: asymmetry Motion: Forehead: none Eye: incomplete closure Mouth: slight movement
VI	Total paralysis	No movement

Although these methods are easy to apply, the resulting informations are subjective and ambiguous, vulnerable to observer disagreement and, therefore, none of these systems provides precise, reliable and quantitative measurements regarding dynamic facial movement. In addition, and generally speaking, the evaluation is made to the entire face

and, being so broadly, these subjective methods do not provide useful information about a specific area for facial reconstruction.

Questionnaires are another category of qualitative methods, aiming to better understand the impact of certain surgical procedures in patient's life, and thereby improve the outcomes of the intervention, comparing the results for an individual subject, across different procedures or even surgeons [17]. The use of these standardized instruments allows the measurement of patient's satisfaction, functionality and quality-of-life and, thus, they are subjective and difficult to interpret, varying from patient to patient, according to its expectation and opinion.

The most used and studied quality-of-life and body-image questionnaires, for being the most appropriate for a general assessment of patient satisfaction, are, among others, the *Multidimensional Body-States Relations Questionnaire* (MBSRQ), *EuroQol* (EQ-5D), *Health Utilities Index* (HUI) and *Derriford Appearance Scale 59* (DAS59) [12, 16].

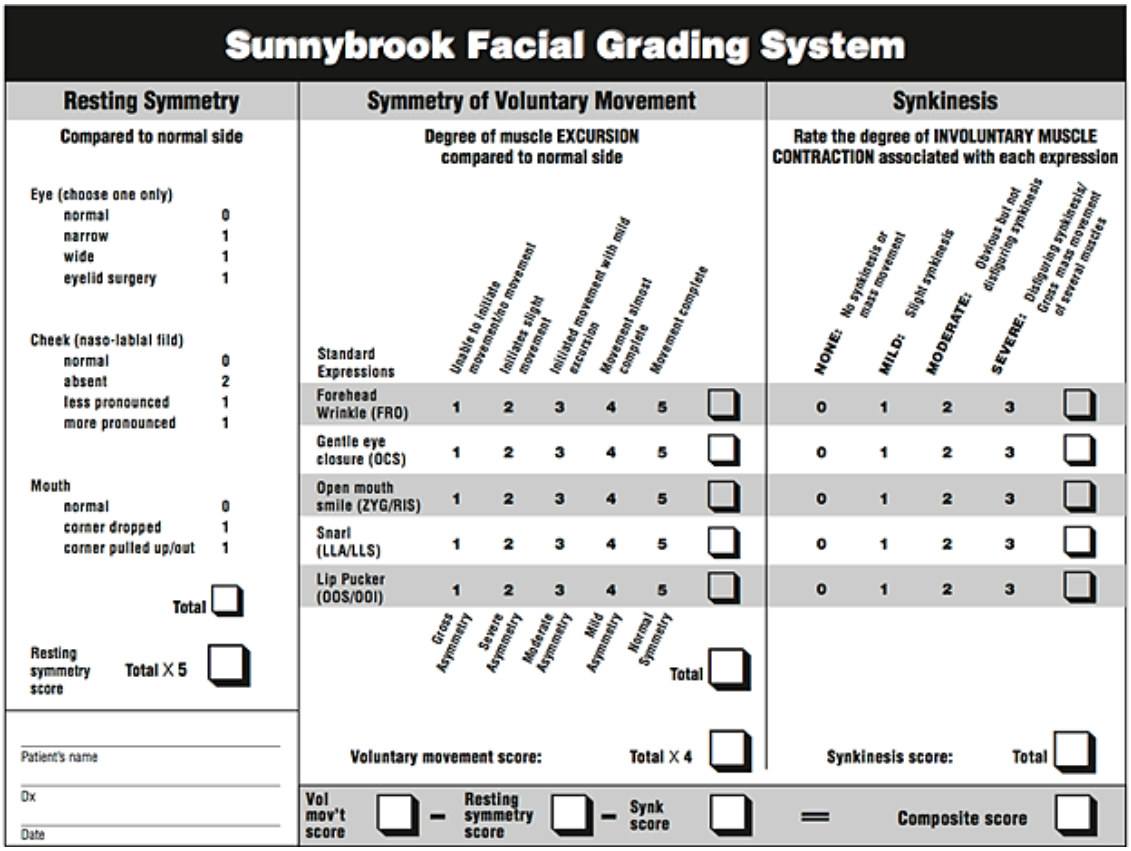
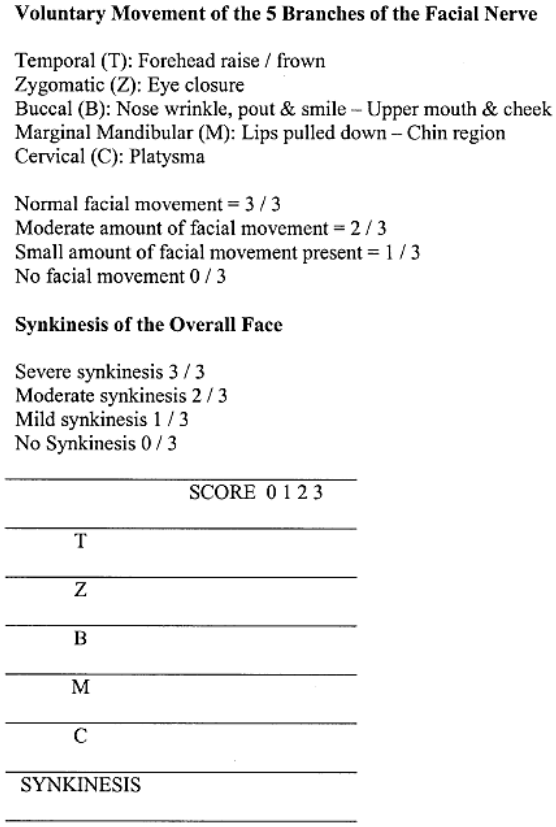


Figure 3. *Sunnybrook Facial Grading System*. Retrieved from [18].



Voluntary Movement of the 5 Branches of the Facial Nerve

Temporal (T): Forehead raise / frown
Zygomatic (Z): Eye closure
Buccal (B): Nose wrinkle, pout & smile – Upper mouth & cheek
Marginal Mandibular (M): Lips pulled down – Chin region
Cervical (C): Platysma

Normal facial movement = 3 / 3
Moderate amount of facial movement = 2 / 3
Small amount of facial movement present = 1 / 3
No facial movement 0 / 3

Synkinesis of the Overall Face

Severe synkinesis 3 / 3
Moderate synkinesis 2 / 3
Mild synkinesis 1 / 3
No Synkinesis 0 / 3

	SCORE 0 1 2 3
T	
Z	
B	
M	
C	
SYNKINESIS	

Figure 4. *Sydney Facial Grading System.* Retrieved from [18].

Having regard to the specific case of the face, some instruments can be used, as is the case of *Rhinoplasty Outcomes Evaluation*, *Facial Lines Treatment Satisfaction Questionnaire*, *Facelift Outcomes Evaluation* or *Blepharoplasty Outcomes Evaluation*, specific for the common procedures of facial plastic surgery [15, 17]. However, all of them are mainly used in cosmetic plastic surgery and therefore are beyond the scope of this study. An extensive description can be found in the literature [12, 15-17].

In addition to the disadvantages mentioned for the grading system methods, which relate to the subjectivity and ambiguity of these tools and their lack of facial movement quantification power, these questionnaires have a drawback which concerns their dependency on patient's opinion, besides the specialist judgment.

Some studies claim the existence of quantitative methods, using the questionnaires and grading systems presented, arguing the existence of numerical scales to rank the degree of movement/paralysis or quality-of-life. However, these methods, such as those presented above, may be considered semi-quantitative, since they do not only have qualitative concerns, presenting numerical rating for each item and final grades based on numerical calculations [17, 18].

Although, as mentioned, there are multiple approaches to subjectively evaluate the results of surgical procedures based on the specialists' and/or the patient's opinions, little progress has been made to effectively quantify these qualitative results in an objective manner.

2.4.2 Quantitative methods

Several studies have tried to address the limitations of the qualitative methods and, so, create a system that allows the quantification of facial movement.

One approach used to objectively determine facial movements is to evaluate the luminance changes in the face. The determination of this luminance allows the creation of subtracted gray-scale images, which are then binarized (see Figure 5), each one corresponding to a movement of the face. This objective scaling method, known as *Objective Scaling of Facial Nerve Function Based on Area Analysis* (OSCAR) [31], although having advantages with regard to the easy interpretation of results, is limited by the large computation time, required for processing the images, and by the need for absolute stability of the patient's body during the measurements, which restricts significantly its applicability in the medical field.

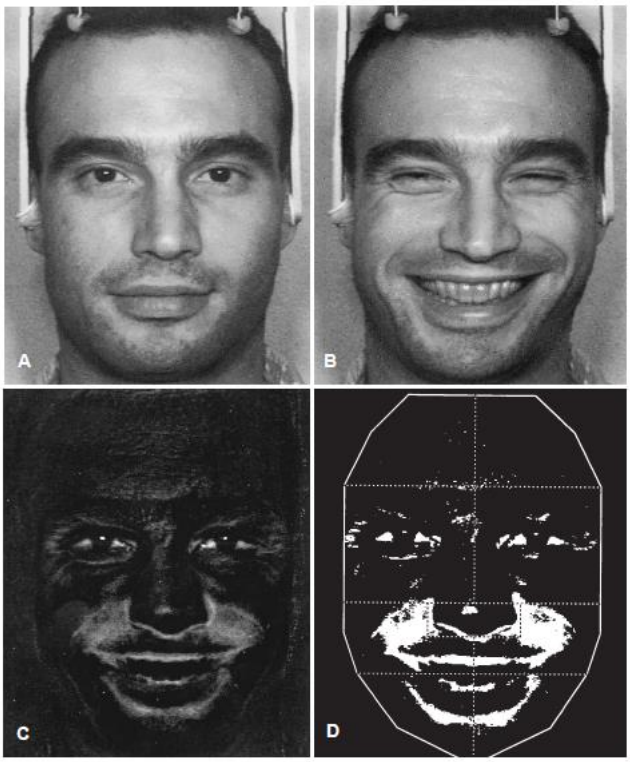


Figure 5. Illustrative image of voluntary movements and its processing. A – Resting phase. B – Motion phase (smiling). C – Subtraction of A and B. D – Binarized image C. Adapted from [31].

In order to measure the movement of the face, facial point systems are possible approaches. The first efforts in this sense were made using manual measurements of distances between fixed points during several facial expressions, with simple measuring instruments [53]. Obviously this rudimentary method fails in the determination of the dynamic motion as well as the direction and angle of points' movement.

Another approach to develop an objective method of measuring surface of the movements was described by Fields and Peckitt [24]. The authors developed a facial nerve function index, which relates, by a percentage, the difference between the distances at rest and during the smile, on the side not affected, and the difference between the same distances but on the normal side of the face. Thus, the mouth expressions were depicted by a facial function, and, in fact, any region of the face could be analyzed based on this

relationship [24]. Although it is a quite simple and objective method, this system does not provide dynamic information about the speed, acceleration or synkinesis, and lacks accuracy.

The first attempts to use cameras to assist these measuring processes were studied using projection of images that represent the position in rest and during the smile, in order to analyze the direction and extent of movement of the upper and lower lips, nasal labial folds and nasal base. Thus, the variable positions of the reference points around the mouth and lower face were measured quantitatively, during the movement [32]. The results of this study were not applied in the medical practice since this technique was time consuming and lacks of dynamic information relevant to specialists.

Image subtraction techniques are another methods used to analyze the facial function and thus calculate its movements. This method operation is based on the premise that, if an initial image of the face at rest is subtracted from following image-frames that depict face movements, it is possible to obtain pixels with different "colors" in a given area and their sum is used to quantifying facial movement.

Neely et. al. [54] reported initial results using this technique, having pioneered the use of digitized images for the analysis of facial movements. As explained, using the subtraction method, the authors obtained an image in which areas of the face that change their gray scale value appear bright. Consequently, and by thresholding, any area of the face that moves is represented in white and the other which do not experience movement turn black. Through this analysis for all video images, it was possible to build a dynamic strength-duration curve. Besides that, the obtained data showed a correlation between the facial movement quantification and the HB system. This system fails to provide information about the magnitude and direction of the facial movement, being insufficient to fully characterize the facial movements. Following publications of the same authors [55] were aimed to upgrade the system by adding a subjective component (that qualifies the degree of movement using a grade from 1 to 4) for specific areas of the face. Thus, for each area is assigned a score and, at the end, these scores are summed, obtaining a composite index for each half of the face. However, this system places more emphasis on subjective assessment, based on these scores, and so leaves the computerized objective grading systems' category.

A study by Wood et. al. [30] aimed to examine the variability of two facial movements (eyebrow elevation and smiling) using an objective method called video micro-scaling. This technique enables the positioning of computer-generated lines, which act as measuring scale, over pre-recorded video, in each movement stage, so that the distance between the individual lines is measured. Although simple, there are shortcomings to this method. The fact that it is necessary that the patient's head remains in a stable position relative to the camera as well as the high computation time required, turn major limitations. In addition, the subjectivity necessary to set the exact required position of the face contour and the lack of information concerning the movement velocity are other weaknesses of the method.

In line with the previous method, a more recent one was studied, which, using commercially available equipment (*Adobe Photoshop* program), performs measurements between video frames that were overlapped during the various stages of facial movement (at rest or smiling) [22]. The measurements on the overlaid image were made using the program tools and the system can be used to evaluate other movements, besides smile movement, such eye closure, nose movements and forehead elevation. Despite its simplicity, and as the previous ones, this method is time consuming and does not provide relevant and dynamic information about the movement.

Using the same computer software, Sargent et al. take advantage of a camera to obtain images of different facial movements, later analyzed by this program [25]. As in [54, 55], image subtraction techniques were used, and the obtained surface was manually outlined for area calculation. The ratio of areas of the normal side of the face (unaffected side) is taken as a comparison index, which can be correlated with the HB. The images were then manually aligned and the relationship between the movement of the marked points on the face and the number of remaining pixels from the subtraction method was studied. Despite the promising results, which concluded that the movement of the oral commissure when compared with the apparent area of face motion had high intersubject variability, some limitations exist, such as the large number of manipulations required, multiple sources of error and observer subjectivity at each stage. In addition, it was not possible to obtain information about the dynamics of facial movement in 3D [25].

A new procedure for simultaneous and multiregional measurements of bilateral facial movements was studied by Johnson et. al. [33] and the new developed method is called *Maximal Static Response Assay* (MSRA). After collection of photographs in resting position and movement, by placing physical markers on facial landmarks, the authors calculated the displacement of marked points using a software program developed for this purpose. The center of these markers was manually detected in repose and maximum facial movement frames and the displacement between these two positions relates to a quantitative measure of motion. Calibration was done using a ruler placed on the patient's face and the determination of the facial movement was possible after overlaying a grid on the images. This system has been used to study the synkinetic motion, depression incidence and the effect of the impairment of facial movement [21]. Although allowing the calculation of the movement in various locations of the face, as well as detect cases of synkinesis given these regional measurements, always compared to the resting state, this system is time consuming for the operator and patient. Besides that, it experiences distortion in measuring the movement along the photographs and no information regarding motion between repose and maximum movement is given, leading to an insufficient representation of the actual path of facial markers.

Similar to the previous one, which used the landmark-based system, another study, more complete, that also aims to describe and quantify facial movement, was presented by Isono et. al. [27]. This facial marking system consisted in 10 points per side (plus 4 in the midline) and the movements were recorded with a video camera, before being computationally analyzed. The distance of displacement was measured using an image-editing software and, instead of using specific points of the upper face to represent the movements, these displacements on each side were summed. In this way, these results have made it possible to create a ratio measure of the average between the movements of the right and left sides, for patients with and without facial paralysis. The fact that this measure may be converted to a graphical representation can provide a way to evaluate the recovery of the facial nerves and, therefore have clinical utility. However, this technique has a large computation time due to the time needed to collect, analyze and represent the data, and does not provide 3D information about the facial movement.

In order to quantify the facial movement in a new and more precise way, another method was developed, by Wachtman et. al. [21], and is called *Automated Face Analysis* (AFA). The AFA was validated by comparing its performance with the quantification system mentioned above, the MSRA. Unlike the latter, the movement of facial features is automatically tracked

in the image sequence through specific software, that uses the dense flow estimation (Figure 6), without the need for physical markers, and the method only requires to manually mark the first frame of the image sequence and then the points are tracked automatically. Besides that, the linear motion is estimated across the images, and not only between the first and maximum movement frame. With this in mind, the representation of the actual path of facial movement was provided, improving the MSRA. However, the system still does not provide 3D information about the facial motion.

Figure 6. Illustrative images of dense flow extraction results. Retrieved from [21].

In another study, a video computer-assisted system, called *The Peak Motus Motion Measurement System*, was used to objectively quantify the displacement on each side of the face, based on preselected markers, whether in normal individuals or with paralysis [28]. The aims of this study consisted in comparing the percentage of displacement in asymmetry between patients with and without facial dysfunction, quantifying the synkinesis in patients with this facial impairment, and determining the correlation between the quantitative measurement results, express as the percent asymmetry, and the subjective HB rating. In order to achieve these goals, and after the researcher select a center of each facial marker during the first frame, the *Peak Motus* software is capable of tracking the points across all frames, by automatically deriving the *x*- and *y*-coordinates for each marker's location. The asymmetry in marker displacement for a given coordinate was calculated taking into consideration the maximum and minimum values for this marker, during the expression movements. The difference between these two values represented the displacement and, finally, the asymmetry was calculated as a ratio between the side-to-side difference in displacements for both sides and the sum of these displacements. When the degree of displacement in a particular facial location exceeded the 95th percentile, with reference to patients without impairments, the presence of synkinesis phenomenon was confirmed. The main advantage of this method is its ability to measure small movements and to extract facial features either spatial or temporal movements, obtaining dynamic information. However, it is also a method with a high computation time, which does not work in a fully automatic manner and the movement representation is only achieve using two coordinate axes [28].

The description of the relationship of distance between the resting state and the movement of relevant landmarks during facial expressions was also studied, this time by Hadlock and Urban [56] and the implemented system uses a software called *Facial Assessment by Computer Evaluation* (FACE). This program is an improvement of a preliminary one, the *Scaled Measurement of Improvement in Lip Excursion* (SMILE) system [57], which primary objective was to quantify the oral commissure movement. The interface of this preliminary system is shown in Figure 7. The latter, through the initial measurement of the white-to-white corneal diameter (measured horizontally trough the pupillary center), allowed the determination of a specific vector, for each photograph, in order to calculate, for the same photograph, the facial dimensions, proportions and angles, only regarding the lip

excursion during smiling. In turn, the FACE software, with a broader application, can measure relevant facial distances at rest and during expressions. Although it has been proven to be more efficient and accurate when compared to measurements using the ruler tool in *Adobe Photoshop*, this software does not provide dynamic motion information, since the measurement is restricted to two movement's stages, not fully characterizing the facial motion.

2.4.2.1 Three-dimensional quantitative methods

Other searches are directed towards the quantification and representation of facial movement in three dimensions. In fact, the extremely complex facial movements can only be completely and adequately described using a 3D analysis [34].

Figure 7. Preliminary SMILE interface (implemented in *MATLAB* software), for measurement of lip excursion during smiling. Retrieved from [57].

One of the first studies that used a 3D measurement system in order to track the movement of pre-defined points on the face was developed by Frey et. al. [34]. This method consisted of a complex mirror system, constructed to provide three different views of the face while on the video screen, a video camera and a specific computer software developed for standardized 3D analysis of the facial movements recorded on video and for data representation. The latter [computer software] is able to calculate three-dimensional distances and movements from the 3 views: frontal image and right and left mirror images. The degree of symmetry can also be quantified, comparing the right- and left-sided distance alterations, between two specific points during the movements. Although such systems are capable of providing dynamic information of the motion, the overall cost reduces the usability of these methods for clinical implementation.

A new objective measurement system was proposed by Hontanilla et. al. [11], in 2008, which aims to address the limitations of the above. This system is called FACIAL CLIMA and consists on a facial motion capture system that utilizes three infrared-light cameras for video recording and computational analysis with STT capture algorithms. The tracking of 3D markers' trajectories, placed on the face of the patients, is done by two simultaneous processes: extraction of 2D coordinates of the projections of the markers in the obtained

images and determination of the 3D coordinates of these markers. The identification of 3D coordinates that belong to a particular marker is made by labeling. Among all the parameters evaluated in this study, which are represented in Figure 8, may be highlighted the area of frontal triangle, the velocity of eye closure, the amplitude of cheek area or even the smile angle. By comparing with the methods presented above, this system has relevant advantages which may allow a fully characterization of face motion. The calibration, points' capture and static and dynamic information processing steps, which for the traditional methods are described as time-consuming, take only a few seconds.

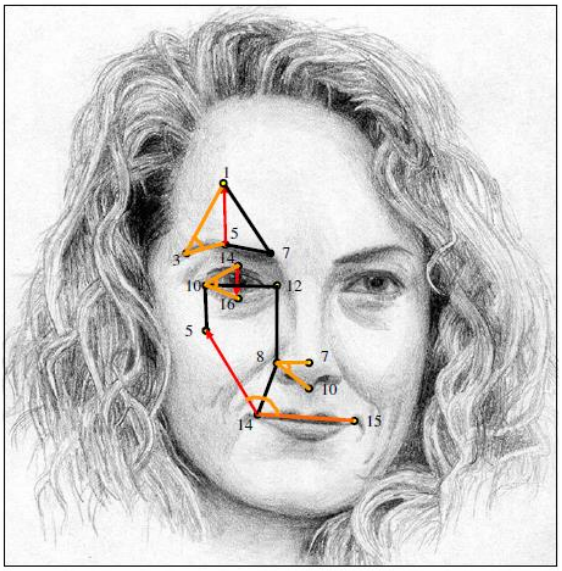


Figure 8. Facial markers used to determine the parameters in the FACIAL CLIMA system. Frontal segment between 1 and 5; Palpebral segment between 14 and 16; Zygomatic segment between 5 and 14. Frontal area among plots 1, 3, 5 and 7; Cheek area among 10, 12, 5, 8 and 14; Frontal angle among 1, 3 and 5; Palpebral angle among 10, 14 and 16; Nasolabial angle among 7, 8 and 10; Smile angle among 5, 14 and 15. Retrieved from [11].

Besides that, FACIAL CLIMA provides information not only of the smiling motion but also of all kinds of facial movements, in a three-dimensional way. The main advantage is even its ability to capture and quantify small movements over the entire face. Along with that, the software includes extras that can be of value for clinical analysis, as is the case of a patient database, biomechanical graphics and a report module that automatically generates reports with the obtained information [11]. Although it seems to be a complete system that can be applied in the medical field, the high cost of the image acquisition system (e.g., the used cameras), and its complex setup are the major limitations for clinical integration.

A new real-time video acquisition system (3D VAS) was developed for the purpose of measuring the positions of objects and the absolute motion of geometric shapes, in real time [35]. In order to evaluate the application of this system for objective measurements of facial movement, Mehta et. al. [26] assessed this method in a healthy group of subjects and patients with facial paralysis, while conducting pre-defined movements. The 3D VAS system operation is schematically shown in Figure 9. By using a digital light processing (DLP) projector and three-step phase-shifting algorithm, the 3D VAS system performs real-time 3D shape acquisition. More specifically, the projector receives from the computer a colour fringe pattern and projects the correspondent three phase-shifted fringe patterns. Afterwards, a high-speed CCD camera, synchronized with the projector, is used to capture the reflected

fringe images and, subsequently, using the algorithm, the 3D shape can be reconstructed based on the three fringe images [35]. Mehta et. al., through clinical trials, proved that this high-resolution, real-time 3D-absolute coordinate measurement system was able to accurately quantify facial movement. However, the authors acknowledge that this study has limitations, which concern the small number of patients in whom the system was tested. Furthermore, this method, besides being costly in terms of required hardware, must be improved in order to allow the automatic collection, analysis, and tracking data throughout the video sequence [26].

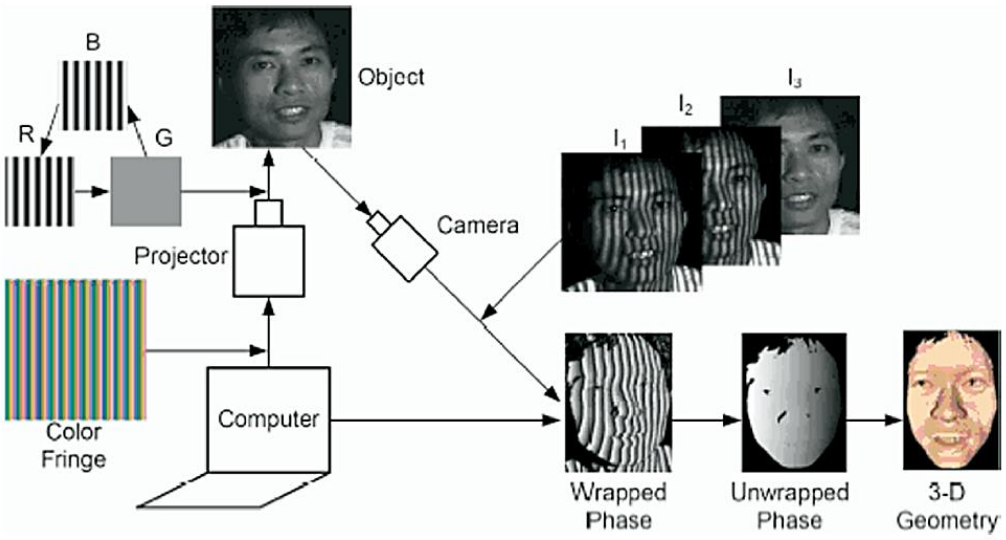


Figure 9. Operational diagram of operation of the 3D VAS system. Retrived from [26].

Finally, other work related to the analysis of facial movement using 3D video systems was conducted by Tzou et. al. [1], with the main goal of assessing the facial paralysis. For this purpose, both sides of the face were compared taking into account the 3D movement of the points between the resting position and motion on each facial expression. This 3D video-analysis system consisted of a set of mirrors, a calibration grid and a commercial camera. In order to observe the movement of each standardized anatomical point that was placed on the subject's face, the patients were pre- and post-operatively videotaped, the data transferred to the computer and analyzed by *Facialis* and *FaciShow* softwares. These were then used to calculate the 3D coordinates of each landmark, which two- and three-dimensional paths were graphically represented. In fact, it was possible to observe the movement or variation in the distance as a function of time, between the landmarks. Although it has been validated and proved consistent in assessing facial palsy, this method has a major limitation regarding its computational time.

Beyond these techniques, which allow obtaining information about the 3D motion from markers located on the face of the patient, other methods have been developed that allow tracking of anatomical points without physical markers. Currently there are two commercially available systems that are capable of performing this type of motion analysis in 3D: the *3dMDfaceTM dynamic system* and the *4D capture system* [58]. The operation of these methods is beyond the scope of this study and a detailed description of each can be found in the literature [58].

2.4.3 Summary

Typically, the evaluation of facial movement is done using questionnaires or grading systems, which seek to assess psychological, social and physical aspects of patients that undergo surgeries. Although easy-to-use, these subjective methods are neither accurate nor suitable for assessing the outcomes of plastic surgery, either to compare between techniques and surgeons, or to choose the best treatment to be carried out. In addition, these methods lack of objectivity, since are based on the opinion of the doctor/patient, and, most of them, are not universal and standardized tools.
Thus, the need has arisen to develop quantitative methods, in order to objectively evaluate the facial movement. Although some of these methods can provide static and dynamic information or even contain 3D information, they have limitations in what concerns their computational and hardware costs, being impractical at the clinical level. An overview of these objective analysis systems developed in the last decades is presented in Appendix A. In summary, the full potential of objective measurement techniques is yet to be realized.

With this in mind, and to overcome these limitations, it is necessary to develop a quantitative method, which provides complete information about 3D movement of facial vectors, while being simple and computationally efficient. Furthermore, the information must be of clinical use, which can be integrated in the medical environment [20].

It should be noted that both qualitative (based on the opinion of the patient and physician) and the quantitative methods are necessary to obtain complete information regarding the effectiveness of plastic surgical techniques and so that the results of such techniques are the most accurate as possible.

2.5 Stages for measuring facial movements

2.5.1 Face landmarking

One of the first stages that need to be performed for the measurement of facial movement is face landmarking. This process is an important step for the following operations, which are focused on the face, such as face description, recognition and tracking expressions, movement determination, etc. It can be defined as the detection and localization of certain characteristic facial points, having a particular geometry, anthropometric features and biological meaning [59, 60]. Actually, in order to confirm the importance of using these landmarks, rather than other points randomly placed on the human face, Gupta et. al. studied this effect of choosing facial fiducial points in the performance of their 3D face recognition algorithm [61, 62]. What the authors did was select particular facial points whose distances were between relevant landmarks, showing that, when the anthropometric distances were replaced by distances between arbitrary points, the performance of the algorithm decreased significantly. Thus, they concluded that the extraction of features and measures are affected by the choice of specific facial points and, so, the anthropometric landmarks are essential for a complete description of the face.

Once landmarks are extracted from faces, its utility stems from the information that can be extrapolated given their position.

Nowadays, in the field of computer vision, it is essential to obtain more complete and dynamic information, derived from 3D models and sophisticated measurements, always using the anatomical landmarks as an intermediate step. However, this landmarking process has

been used for over a century. In fact, in the past years, the craniofacial landmark identification practice has been essential in order to perform anthropometric measurements of cranial variation in different races and ages, or even to study the facial morphology, such the analysis of facial abnormalities, growth changes, etc. [63]. One of the first studies which portrays the use of facial landmarks was performed by Farkas [64]. In this seminal work in the craniofacial anthropometry field, a database of anthropometric standards was constructed by measuring and comparing multiple features (whether linear, angular, or based on contours) and proportions, in a large population sample (over hundred people). In addition, Farkas used anatomical landmarks with the aim of describing the best methods for facial measurements and included, for the purpose, 47 landmark points to describe the face [64, 65].

The anatomical landmarks can be divided into two major groups. The soft-tissue landmarks are on the skin and are easily identified on 3D point clouds generated by scanning or on images. In turn, the hard-tissue landmarks are located on the skeleton and are only identified through radiographs [59]. In this sense, they are rarely used and, regarding the aims of this study, they will not be considered. Indeed, depending on the application they are intended to be used, these soft-tissue points may vary in number. In Figure 10 are shown the soft-tissue landmarks most commonly used in the literature. It should be noted that these landmarks can be further divided in 2 other groups: primary landmarks (which are easy to identify and are usually used for the purpose of facial tracking and detection) and the secondary or auxiliary ones (which are more difficult to detect; in fact their detection is guided by the primary landmarks, and are usually associated with facial expressions) [60].

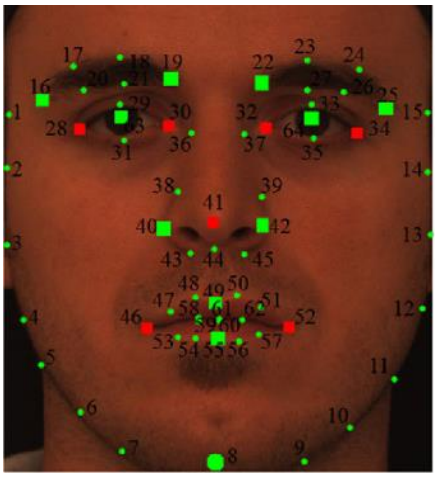


Figure 10. Soft-tissue landmarks. The squares represent the primary landmarks: in red, the most fiducial ones. Green dots indicate the secondary landmarks, totally 64 facial landmark points. Secondary landmarks: 1 – Left temple; 8 – Chin tip; 2-7, 9-4 – Cheek contours; 15 – Right temple; 16-19 – Left eyebrow contours; 22-25 – Right eyebrow corners; 29, 33 – Upper eyelid centers; 31, 35 – Lower eyelid centers; 36, 37 – Nose saddles; 40, 42 – Nose peaks (nostrils); 38-40, 42-45 – Nose contours; 47-51, 53-62 – Mouth contours. Primary landmarks: 16 – Left eyebrow outer corner; 19 – Left eyebrow inner corner; 22 – Right eyebrow inner corner; 25 - Right eyebrow outer corner; 28 – Left eye outer corner; 30 - Left eye inner corner; 32 – Right eye inner corner; 34 – Right eye outer corner; 41 – Nose tip; 46 – Left mouth corner; 52 – Right mouth corner; 63, 64 – Eye centers. Retrieved from [60].

There are also morphometric measurements and geometric features that can be extracted from the landmarks. The study, developed by Vezzetti and Marcolin [59], provides an extensive description of the various types of morphometric measurements that can be

obtained using facial landmarks. Among all of them, the authors concluded that the *Euclidean* and *geodesic* distances are the most commonly used in methods for face evaluation. Depending on the purpose of each study, these measurements can be combined to obtain the best possible results.

The face landmarking process can be done manually, by selecting specific points associated or not with markers placed on the subject's face that act as fiducial points, or automatically, through specific algorithms in order to detect the points without requiring physical markers. This resent study uses the first approach, which will be described later. This system, called marker-based tracking, requires the placement of multiple landmark identification on the patient's face and, subsequently, these markers are captured by a camera or recorded video. The physical markers can be of two types: active and passive. Systems based on active markers use points that are activated by an infra-red signal and respond to a corresponding signal. In turn, the passive markers reflect light back, since are coated with retro reflective materials [58]. Although the marker-based tracking systems are a simplest implementation, some studies indicate that these physical markers may inhibit the natural and spontaneous facial movements, thereby limiting the analysis of pre-selected features [1, 21]. To avoid this limitation, an alternative methodology is to use ink dots as markers which, while generating the necessary fiducial points to the tracking system, are totally imperceptible to the subject and do not affect facial movements.

2.5.2 Tracking of facial features

After the detection/selection of certain facial landmarks/points, tracking these points of interest allows the analysis of their motion, an essential step to quantitatively characterize the facial movements, in a complete way.

The facial feature tracking consists, then, in detecting and following facial features points (facial landmarks), surrounding facial components, over multiple frames [66, 67]. Once a feature is tracked, it becomes a set of coordinates representing the position of a specific point in the image, across a series of frames, and all features' tracks that were created can be used immediately for motion tracking or to calculate trajectories' information.

In section 2.4.2, apart from the methods that have been mentioned, namely [1, 11, 21, 28], which required to perform methods of facial features tracking in order to able to quantify the facial movement, the tracking of facial points can also be used for other purposes, such as face detection and registration, expression recognition, head poses, facial animation, etc. [66-69]. The existing methods can then be divided into two categories: appearance based methods and feature based methods. The first approach uses generative linear face models to capture the shape and texture variations of faces. Such methods include *2D Active Appearance Models* (AAM) and *3D Morphable Models*, the most commonly used. In turn, the feature based approach tracks local facial features, using an aggregation of these features to represent the face. In this case, this method can use *Active Shape Models* (ASM) or other features for tracking. Although presenting better generalization capability, they may lose quality in tracking due to the lack of semantic features and occlusions when the face is in different orientations [70-72].

Some authors further agree that facial feature tracking methods can be differently classified, in another two categories: model-free, that uses point trackers without prior knowledge of the object/face, with feature points being tracked individually by performing a

local search in the following images, and model-based tracking algorithms, that aim to model the shape of object/faces, using models, as those previously mentioned [67]. The former have the disadvantage of being susceptible to tracking errors resulting from noise, occlusions, etc. In turn, the model-based approaches, despite being of greater complexity, showed better results for feature point tracking, in situations of large and disparate movements [73].

Thus, researches on the methods that take advantage of facial features tracking for a wide variety of applications, whether or not based on models, may be powerful tools to choose the most suitable method to perform the tracking of facial points and, consequently, to obtain a quantitative analysis of their movement across the image's sequence. Hereupon, recent studies with the aforementioned characteristics will be presented in the next section.

2.5.2.1 Current methodologies for facial features tracking

In a study conducted in 2010, a *Regularized Maximum Likelihood Deformable Model Fitting* (DMF) algorithm was developed with the aim of performing face tracking, using depth cameras [70]. For this purpose, it was necessary to track the facial movement in subsequent frames, by using feature points in texture images. These points were matched across frames and integrated into the DMF framework. The point-to-point method was used to ensure proper tracking of the facial expressions. The tracking process used the cross correlation method, by matching the feature points of the current frame with the ones in the previous frame. The proposed algorithms have limitations that were detected when the performance was tested, in situations of fast and large movements.

In order to accomplish automatic tracking of facial landmarks for the recognition and analysis of emotion's dynamic, a study that presents a feature-based framework using multi-kernel learning was developed [69]. The *Advanced Multi-kernels* algorithms were then applied in order to perform the tracking in the same way that the traditional features-based tracking methods are used to do, i.e., track each point by searching each image area, chosen because of its similar appearance to the one in the first frame, for the best matching position (see overview in Figure 11). However, with the purpose of improving the detection of landmark points in the following frames, and therefore to avoid the characteristic drift of other features-based tracking methods, prior knowledge has been incorporated in the appearance of facial landmark in the current frame.

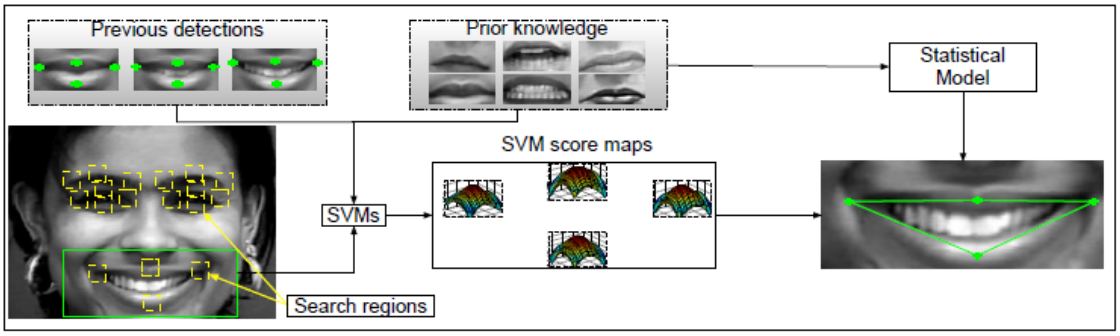


Figure 11. Overview of the system that uses a multi-kernel learning to perform tracking. Retrieved from [69].

An alternative way to track facial features is the use of *Shape-Constrained Linear Multiresolution-Selected* predictors, investigated in the study by Eng-Jon Ong et. al. [74]. This linear predictor (LP) method is then used to track features, in which each LP provides a

mapping from sparse template differences to the displacement vector of a tracked feature. Each is characterized by five components: 2D location of the feature, linear mapping to the displacement vector, 2D offsets for support pixels positioning, a vector of base support pixel values, whose dimensions follow the number of these pixels, and a bias factor to the mapping. By grouping multiple LPs into rigid flocks, it is possible to individually track a feature point over multiple frames. Its performance was compared with methods mentioned in the previous section (the AAM and ASM), and proved to be more accurate than the first one, with less complexity and smaller training size. However, this method is dependent-person and susceptible to occlusions.

Indeed the ASM and AAM methods are commonly used approaches for facial features tracking. The former, originally proposed by Cootes et al. [75] is a statistical model-based approach, aiming the representation of deformable objects, that define sets of feature points' shapes. These facial shapes are generated and iteratively deformed so as to fit on the face in a given image and to maximize the fitting of the information present in those frames. Since the positions of feature points are simultaneously updated, the interaction between these points is interdependent. The fact that this method is person-independent reflects a lower accuracy. Furthermore, it requires a large initial training set and does not allow the individual feature tracking [67, 74]. The use of person-specific trackers invalidates the need for large training data sets, which prompted the development of AAMs, initially projected by Cootes et al. [76]. The aforementioned limitation is, then, overcome through shape-constrained and face appearance information during the tracking stages. This tracking method has, nevertheless, restrictions, stemming from the fact that individually feature tracking cannot be performed and needs the entire face shape and texture for efficient feature tracking [74].

Unlike the method of Eng-Jon Ong et. al. [74], that used only intensity information for a real-time tracking, a new study was developed to integrate this information with depth contents. In this way, this approach, by Baltrušaitis et. al. [77], uses a *3D Constrained Local Model*, together with depth information (CLM-Z) to facial features tracking for different head positions. Thus, this new method seeks to integrate the rigid and non-rigid tracking, in order to combine head pose estimation with facial points tracking. These CLM approaches model feature points using the same *Point Distribution Model* (PDM) as the previously mentioned model-based approaches. The PDM is a model for representing the mean geometry of a specific shape and consists of a non-rigid outline and rigid global transformation parameters. After the training images are landmarking and the model is trained on these labeled examples using the *Generalized Procustes Analysis* and the *Principal Component Analysis* (PCA), the fitting process is performed in order to estimate rigid and non-rigid parameters, whose variation within limits learnt from the training set, leads to the generation of new shapes [78]. After being compared to available datasets with existing ground truth head pose data, this innovative method showed better performance in terms of accuracy, either in a single image as a video sequence. However, since it has not been tested in external noise conditions (only clean training data), the noise influence was not explore and it could affect the algorithms' performance [77].

Still using depth information, combined with color information, a framework was developed by Yang et. al. [71] aiming the face shapes tracking. A *Microsoft Kinect* camera was used to capture low-resolution depth images, in order to estimate the orientation of the head and generate extra constraints at the boundary face, for the selection of the closest

pose. After the location of landmarks, these were tracked and their position is further updated using the global shape constraints. In order to minimize the computational cost caused by ASM in each frame, the individual landmark tracking was done using the *Kanade-Lucas-Tomasi* (KLT) feature tracker, over consecutive frames. Figure 12 shows the system overview. This method is a feature tracking approach that uses the spatial intensity information for searching the position that yields the best match. With this in mind, the registration between two local features is made, by computing the displacement of the feature, and the new location of this landmark is found in the next frame [71, 79]. This technique is faster than the traditional ones since the examination is made in fewer potential matches, between the images.

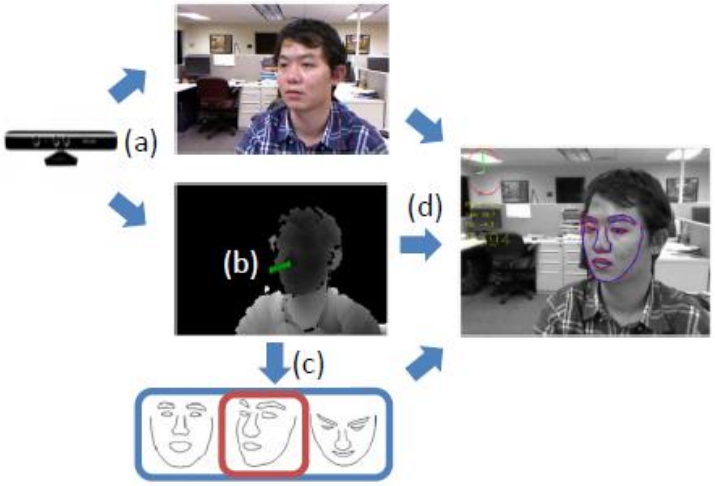


Figure 12. System description: a) acquisition of RGB and depth data using a Kinect sensor; b) Estimation of face orientation based on the depth images; c) Selection of the face subspace of the closest pose to limit face shape; d) Tracking of the face. Retrieved from [71].

A method that combines three types of features, which, individually, discriminate the distortion of the face in different views, was proposed by Zhang et. al. [72]. Taking advantage of the integration of different types of features, it is possible to extract more discriminative information from features correspondences and, consequently, from the face deformation. These three different features are: semantic features, which are equivalent to those defined in ASM and provide correlation between the 3D models and main facial features; silhouette features, that are dynamically matched in different positions, giving relevant information about the facial shape; online tracking features, which are obtained by matching the interest points of the image, which, therefore, can vary over several frames. In this way, a framework for 3D tracking of non-rigid face deformation was proposed, which flow diagram is shown in Figure 13. The tracking of the interest points is made after the selection of these points from the previous frame and some key-frames, matching them to the current frame. The initial 3D pose estimate can be obtained by knowing the 2D coordinates of the interest points from the previous frame and key-frames. Since the deformation parameters and 3D pose of the previous frame and key-frames are known, it is possible to calculate the 3D interest point's correspondences. Although being effective and having low computational cost, this method has the same drawbacks mentioned for ASM technique.

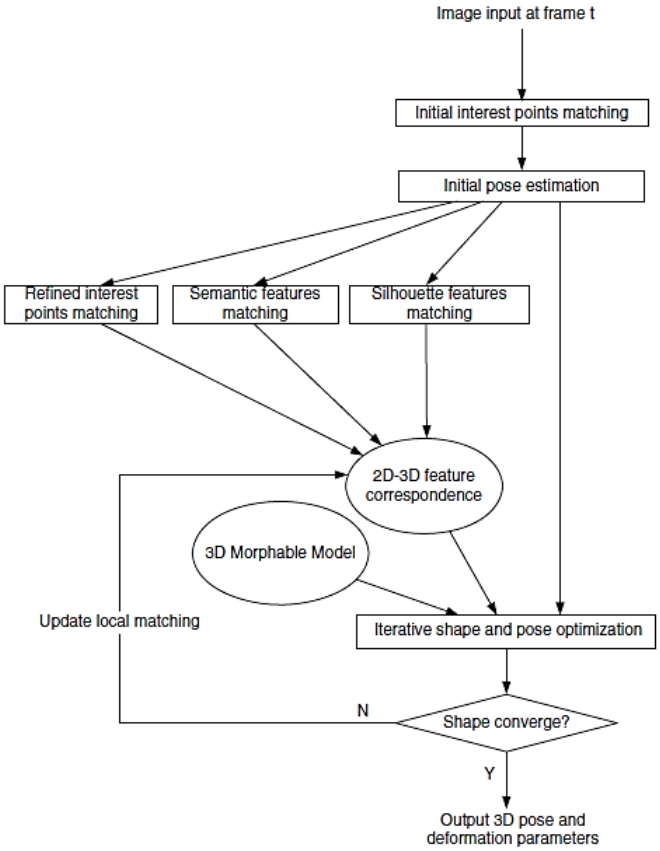


Figure 13. Flow diagram of the 3D deformable face tracking algorithm. Retrieved from [72].

Another study was developed, also using a method for the tracking of facial features, in which the authors tried to solve a general problem that is the elimination of rigid motion of the head, when estimating the facial deformation [80]. This rigid motion effect occurs as the head moves while the face deforms, an undesirable effect if it is intended to accurately measure the deformation of the face. In addition to the elimination of this effect, the authors sought to understand how the facial feature tracking across image sequences could be optimized in order to improve the results. The used feature points are *Scale-Invariant Feature Transform* (SIFT) keypoints, whose descriptor vectors are manipulated to determine the spatial and temporal correspondences across different frames. If no temporal matches are found, the tracking is done using the *Least Squares Matching* method. In the latter, the correlation between the template images and each search image is made, and a template images' alteration can occur during the process. By doing so, this new proposed method is based on simultaneous matching of multiple spatial and temporal template images with each search image, being the spatial template images from the same frame as the searched ones, which is an advantage over the use of fixed templates. Figure 14 shows the different steps of the proposed algorithm, from camera calibration to visualization of face deformation.

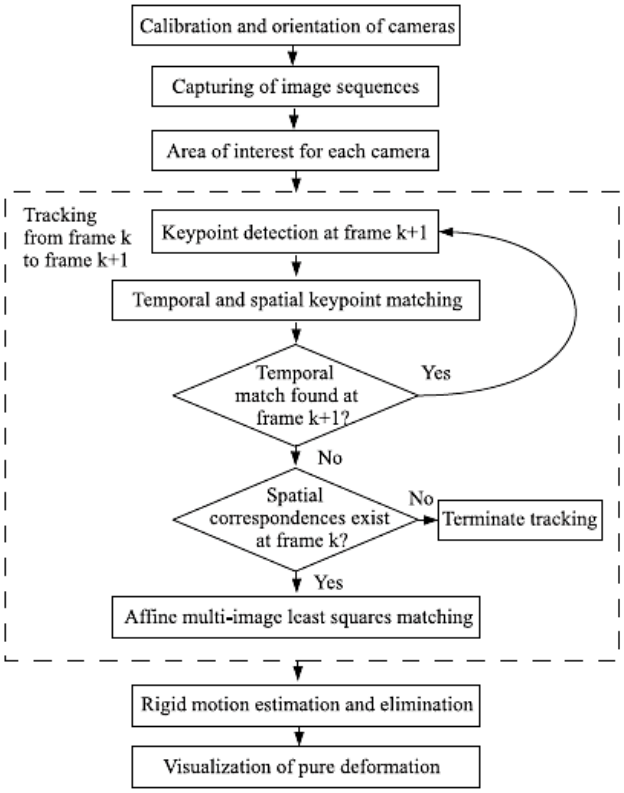


Figure 14. Flow diagram showing the different steps of the proposed algorithm. Retrieved from [80].

Two different studies, both from 2013, were developed for the same purpose, the study the human facial expressions, based on the tracking of landmark points [66, 67]. The first, proposed by Tie et. al. [66], uses *Multiple Differential Evolution-Markov Chain* (DE-CM) particle filters to perform the facial landmarks' tracking along video sequences. This algorithm, developed by Ter Braak in 2006 [81], allows for *Bayesian* computations by *Markov chain Monte Carlo*. Along with this, the feature points were tracked using prior knowledge on the facial feature configurations. In order to find the detection likelihood by maximizing a criterion of similarity between search and candidate points, a kernel approach correlation analysis was developed.

The other one, developed by Li et. al. [67], uses a hierarchical framework based on *Dynamic Bayesian Network* (DBN) for tracking of facial features along with the recognition of facial expressions. The corresponding flow diagram of the algorithm is shown in Figure 15. Initially, the tracking of facial features is performed, using the aforementioned ASM technique. Subsequently, the DBN model is built so that, besides the tracking results can be used for determining facial expressions, this expression recognition can also improve the performance of tracking (two-way interaction model). The model is then trained using training data and subjective prior knowledge.

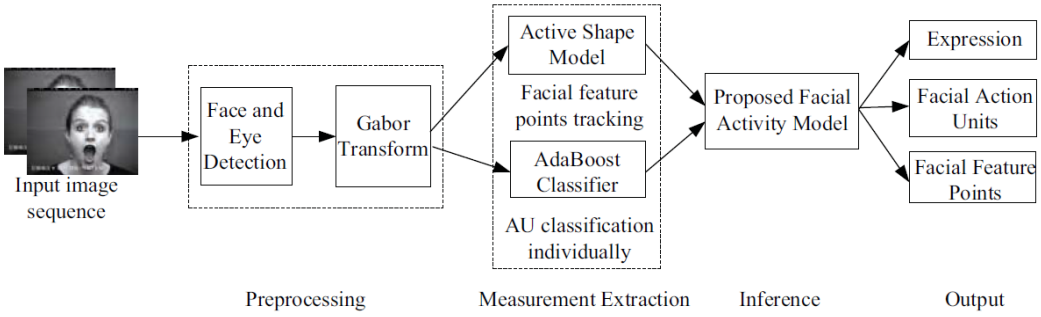


Figure 15. Flow diagram of the facial feature tracking and facial expression recognition algorithm. Retrieved from [67].

Both methods have limitations regarding the performance of their algorithms under specific situations that have not been tested. In the first case, and since the proposed method was not designed to cope with occlusion situations, it is unknown whether its performance will be affected by this feature. In the second study, the model does not handle with image sequences with rigid head movements, and therefore the algorithm may not operate in the correct way regarding these cases of multi view faces [66, 67].

IntraFace is an application that is able to detect and track 66 facial features landmarks in a video, in addition to many other functions related to facial movements. The tracking algorithm is based on a method called *Supervised Descent Method* (SDM), whose aim is to learn a series of parameter updates that minimize the mean of all *Non-linear Least Squares* (NLS) functions, using training data. This data consists in a set of functions where the minima are known. Therefore, SDM learns the descent directions in a supervised way. More information about this particular method is available on [82]. The use of this method was tested for facial feature tracking and the results were very promising. The tracking was performed using SDM for detection in each frame image, initializing the frame with the landmark estimation from the previous frame. The model was trained with 66 landmarks in different datasets (containing facial movements from subjects apparently without pathologies) and the results showed that SDM algorithm never lost track, even in occlusion cases. Although performance results were truly encouraging, a big disadvantage of this method is the fact that the tracking is done only for specific anatomic landmarks, pre-defined and fixed.

Another technique that can be used to make facial points' tracking across a sequence of frames and, therefore, perform motion estimation is called Block Matching algorithm (BMA). This method is based on the assumption that if a specific pattern, at a given frame of a video sequence, moves, corresponding objects will appear in the consecutive frames. The determination of motion vectors that describes the movement or transformation from the previous to the current 2D frame is called motion estimation. In the BMA, the motion estimation is performed by comparing each block, in which the current image was divided, with corresponding blocks and their neighbours in the previous frame, so that motion vectors containing the information about the movement of the block from one location to another, are created. The set of vectors representing the motion of all blocks in the frame represent the motion estimated in this current image. The operation of this algorithm is represented, in a schematic way, in Figure 16.

Some methods use this algorithm in order to perform the tracking of facial features with the most varied purposes, similarly to the methods above.

One of the first methods that used the BMA to achieve this goal was developed by Sobottka and Pitas in 1997 [83], and later in 1998 [84]. These studies describe a method for face segmentation, and extraction and tracking of facial features. The face contour determination and tracking of features tasks are performed using deformable models and block matching, respectively. In this last one, and once the feature points are detected, these can be tracked over frames in accordance to the algorithm operation described above: frame division into initial blocks that can be tracked by finding corresponding blocks in next frames.

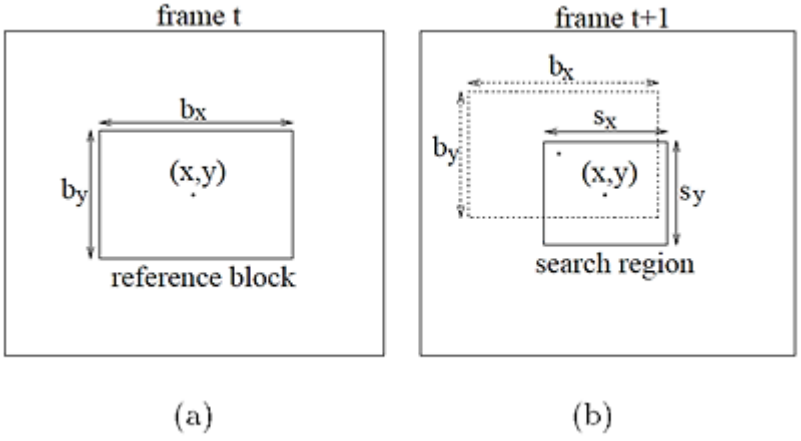


Figure 16. *Block Matching* algorithm. (a) reference block in frame *t*, and (b) search and reference blocks, in frame *t+1*. Adapted from [83, 84].

With a different purpose, i.e., performing face tracking, another study, developed by Valente and Dugelay [68], proposed a method which uses visual feedback loop, i.e. a wireframe face model and a specific and texture are constructed, using 2D patterns synthesized by the face model itself. The use of BMA aims to make the match between real images and synthesized patterns and avoid background areas during this process. The key-feature of this model is the use of *Kalman* filters, which, in this particular case, were used to predict the positions of the feature points to be integrated in the BMA. More information about this type of filter can be found in [85].

The use of the BMA was also found in a study that aimed to reconstruct facial images. The principle of the method developed by Koyuncu et. al. [86] is based on the fact that it is possible to define facial expressions using motion vectors and, thereby, the reconstruction of the original image can be ensured from the image with a specific facial expression. As mentioned, through the BMA, it is possible to estimate motion vectors and, in this study, two renowned BMA techniques were used: *Full Search Algorithm* and *Three Step Search Algorithm*, wherein their performance was evaluated. In fact, the results revealed that both techniques were successful regarding the facial reconstruction and the first showed best results, with a greater percentage of matching. For noisy data, it is still necessary to perform further tests to assess the performance of the proposed method in this particular situation.

A more recent study used the BMA to track the position of the facial markers, manually place in the patient's face, in a sequence of frames [5]. In the tracking process, assisted by the BMA, the position of the various markers was extracted over all frames as a function of

time. The results were then used for spatial-temporal analysis and quantification of facial movement.

Finally, a modification of the original BMA was proposed by Kin et. al. [87], in order to improve the efficiency and performance of the same in tracking facial features. The modification was made to the window shape, which now considers the characteristics of the feature points and the scale/angle changes of the face. Thus, the new window, called radial line window, consists of a set of pixels in the eight radial lines and the results proved that this window shape is more efficient than the conventional rectangular window, since it allows obtaining different measures for tracking individual feature points, at a lower computational cost.

2.5.3 Summary

The analysis of facial movement is of paramount importance in order to create a method capable of objectively and quantitatively assessing the facial movement, with the ultimate goal of supporting and evaluating the reconstructive plastic surgery as well as enabling the comparison between different surgical techniques and evaluation of their effectiveness. The facial landmarking and feature tracking stages are necessary to capture the facial movement, through the trajectories of specific anatomical points during rest and facial expressions.

The facial landmarking methodology is of utmost importance and, in the fact, the choice of specific anatomical locations have influence on the final tracking results. With this in mind, this choice should include well-known and previously studied anatomical landmarks, in order to maximize the stages of detection and tracking, during the facial expressions.

In addition to facial landmarking, there are already numerous studies that take advantage of facial feature tracking methods for different purposes. Most of them have been partly successful in specific situations, so they may have a useful application for tracking of facial features. Others still have limitations related to the efficiency and performance in certain conditions that need to be improved. A particular method was highlighted, the *Block Matching* algorithm, which allows the tracking of feature points and the estimation of motion vectors, across several frames in a video sequence. Besides being a simple technique, the BMA has low computational cost and can be used to track an individual feature point. However, compared to other methods, it has lower robustness in performing more complex tasks. In this way, this technique was the chosen one to perform tracking of physical markers for quantification of facial movement.

2.6 *Depth* cameras in face analysis

The analysis of human motion and tracking has become a major application in computer vision. In order to accomplish this, digital cameras are used, with range devices playing a very important role. The latter, also referred as depth cameras, produce images whose pixels express the distance between a known reference plane and a visible and specific point in the scene. In fact, the use of methods based on the imaging range concept, which is related to a set of techniques that are used to produce this type of 2D images, has steadily increasing as well as the development and production of image corresponding sensors used for this purpose [88, 89]. Range cameras, concerning the sensor device that is used to produce this type of images, may have different operating modes, for example, stereo

triangulation, structured light, and *time-of-flight* (ToF). If the sensing system combines RGB color information with per-pixel depth information, it is also called RGB-D camera.

The concept of stereo triangulation is used by stereo cameras in order to determine the depth determination to points in the scene. These cameras are widely used in computer vision applications and have the advantages of not requiring special lighting conditions, capture data in real-time and also integrate textural information [90]. However, apart from requiring calibration, these cameras have to solve the correspondence problem between different images and depend on surface features to determine depth [90].

In turn, the structured light general principle reflects the process of projecting a known pattern of pixels onto a scene and inferring depth from the deformation of that pattern. An example of this type of cameras launched on the market in 2010 is the *Microsoft Kinect v1* sensor. This motion sensing input device combines a RGB camera along with a depth sensor (an infrared (IR) laser-based projector and an IR camera) and a four microphone array, which altogether allow full-body 3D motion capture, facial and gesture recognition (see Figure 17).

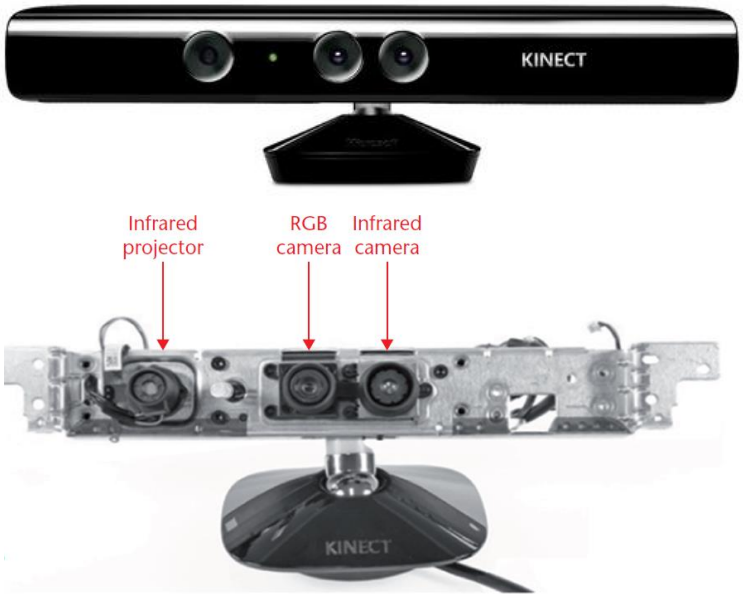


Figure 17. *Microsoft Kinect v1* sensor: external appearance (above) and system's components (below). Adapted from [91].

Based on the previous principle, the *Kinect*'s IR projector sends out a fixed pattern of light and dark speckles, which is captured by the IR camera and compared part-by-part to reference patterns. The variation of the projected pattern against the known one for a fixed distance provides a method to reconstruct the depth map [92]. The default RGB video stream has 8-bit VGA resolution (640×480 pixels), while the monochrome depth sensor has an 11-bit resolution of 320x240 pixels. The depth range of *Kinect* is 0.8 to 4.0 meters, and it is able to resolve close to 1mm in depth coordinates. Both video outputs work at 30 frames per second [91, 93]. Interestingly, it was originally created for game purposes on *Xbox 360* video game console and *Windows* computers, but quickly spread to technological and scientific applications. In fact, some studies have shown that the *Microsoft Kinect* is an accurate device for clinical purposes [94, 95]. Besides that, *Microsoft* has developed a *Software Development*

Kit (SDK) that includes a set of libraries, device interfaces and resource materials to allow the development of application on a variety of platforms [91, 93].

A study by Li et. al. [96] was performed, aiming to compare the performance of some common 3D acquisition devices and to study the feasibility of using the *Kinect* depth data for face recognition under varying pose, expressions, illumination and disguise. After the comparative analysis, it was established that the *Kinect* is a low cost, high speed and compact size sensor, compared to other high-quality devices, such as *Minolta*, by *Minolta Co., Ltd.* or *SwissRanger*, originally by the *Centre Suisse d'Electronique et Microtechnique, S.A.*. However, *Kinect*'s 3D data (depth data) is very noisy and of low resolution. Still, after tested its performance in challenging situations, tasks such as face recognition can be well performed with these low-cost high-speed 3D sensors [96].

Interestingly, sensor systems that provide depth data have been custom-built for years; however at an extremely high price. The RGD-D cameras are now available at low cost due to the development of the *Kinect v1* model.

With the purpose of circumvent certain limitations of the already existing range cameras, a new type of image sensor has been created, called ToF cameras, which integrates the acquisition of accurate intensity (color data) and range information, in real-time, into a single and compact device at a low cost. More specifically, this system, by using a laser or light pulse, determines the distance measuring the signal time-of-flight between the camera and the subject, for each point of the image. Among the several advantages can be highlighted the reduction in shadow areas, efficiency under different lightning conditions and speed. However, interferences and multiple reflections can influence the performance of this system, by the fact that they light up the whole scene [88, 97, 98]. The latest model of *Microsoft Kinect* (*Microsoft Kinect v2*) incorporates this type of camera, with higher resolution and performance than the previous version.

Recently, researchers have developed face analysis-based systems that take advantage of different ToF cameras [70, 88, 90, 97, 99, 100]. In this way, some efforts has been applied to face detection [97], facial feature tracking [88] and head-pose tracking [90], all of them using the *MESA SR3000* ToF camera, and also to perform head tracking [99] and facial expression recognition [100], with the *Canesta* and *ZCAM*™ ToF imaging systems, respectively. Although the previous studies have achieved the purpose for which were earmarked, they all had limitations, as noisy data during fast motion, high hardware's prices and great computational cost, in capturing depth data.

As a result, and for the reasons stated above, the *Kinect* system has a better quality/price ratio. Therefore, in the following section, special attention will be paid to the state of the art that exploits this 3D motion sensing device, in particular the *Microsoft Kinect v1*. Given the availability of this system and the more mature state of its SDK, the *Kinect v1* was the camera of choice in this work. However, except for the drivers used in the clinical data acquisition module, all algorithms developed in the present research are compatible with *Kinect v2* or any other RGB-D camera (that allow different resolution/bit-depth/frame rate).

2.6.1 *Microsoft Kinect v1* in facial analysis context

In addition to the studies developed by Yang et. al [71] and Horta et al. [5], mentioned in section 2.5.2.1, which sought to perform, respectively, face tracking and analysis of the face

excursion, other recent studies have developed methods using the same image acquisition tool – the *Microsoft Kinect v1* system.

In order to detect facial expressions, Vineetha. et al. used the *Microsoft Kinect v1* sensor to obtain the depth information, needed to build and train the model, which is based on *Artificial Neural Network* [93]. Then, a 3D wire frame model of the face was created and facial features from this could be extracted. In this baseline study, the matching process between different facial gestures revealed promising results; although, the network was only trained with three facial gestures and for this reason the system must be improved so as to be able to precisely characterize all face movements.

Facial expressions were again the subject of detailed study but this time in one conducted in 2013 by Seddik et. al. [101]. Again, using the *Kinect* system, both depth and RGB information were used to animate a 3D facial model. The facial expressions were identified on the RGB images of the face and the reconstruction was performed using the filtered depth data.

Finally, human faces tracking and estimation of the face pose were performed, even in different lighting conditions [102]. The face and feature points tracking were made by combining an adaptive correlation filter and the *Viola-Jones* object detection. In this method, the depth information extracted from the *Kinect* camera was used to estimate the face size and to improve the feature's tracking performance. Although accurately working, the authors propose an increasing of the speed of the algorithm as a necessary and further improvement.

2.6.2 Summary

Depth cameras are essential to fully characterize facial movements, given the simultaneous acquisition of RGB-D images. A recent system differs from the other depth cameras in its low price but high performance and speed, combined in a compact device known as *Microsoft Kinect v1* and *v2*. Moreover, its effectiveness has been tested for several applications in the clinical field, even despite its lower quality depth data, when compared with other, more expensive, depth cameras.

In this direction, the use of *Microsoft Kinect v1* in facial movement analysis context has been deeply explored. Several researches considered its potential application to detect and track faces and recognize facial expressions, as well as to estimate the head pose and facial motion. To the best of author's knowledge, little progress has been made on using such cameras to quantify the movement of facial features, so useful to evaluate the performance of surgical techniques and compare clinical methods, and, in fact, efforts should be accordingly made.

Although many of these studies only report preliminary results, with lack of performance and low population sample, they can drive many others in this area, so that, in the future, a complete characterization of the face, as well as quantification of its movement and recognition expressions, may be possible, in an automatically and accurately way. To facilitate the integration of such methods in the medical field, the use of low-cost devices is required, combined with good performance and, at the same time, quick data acquisition and low computational cost. Therefore, the use of *Kinect* technology appears as an interesting option to be used in this context and, for that reason, was the type of camera chosen to obtain the results of the present work.

2.7 Clinical tool to quantitatively assess facial movements: *facegram*

As mentioned in section 1.1, the non-universality and consensually of the existing assessment methods for the outcomes of reconstructive plastic surgical leads to the necessity of creating a measurement system, able to quantify, objectively, the facial movement characterized by static and dynamic features. Thus, it is necessary to create an accurate, objective and consensual system, which gather together a set of morphological, static and dynamic measurements, in order to enable quantitative evaluation of the results of surgical interventions, with an easy-to-learn language and, manly, clinical usefulness.

This new measurement system, proposed to be name *facegram*, should allow the analysis of the trajectory of facial anatomical points, whilst quantifying this movement, displaying all of this in a user-friendly interface for specialized physicians.

Unlike research areas such as face detection, facial features tracking and recognition expression, which have been deeply studied and many solutions have appeared to deal with the corresponding limitations, few studies have been performed to address the spatial-temporal analysis and quantification of facial movement, through measurement of static and dynamic features, and those that were proposed are quite simple and lack some objectivity and standardization.

An early attempt to create a measurement system for facial movements was proposed by Wachtman et. al. [21], and in fact it was previously presented in Section 2.4.2. By using the *Automated Face Analysis* (AFA) method, the feature points tracking was performed and the facial movement was analyzed during, e.g., smile (Figure 18). With these results, it was possible to build a chart, where the movement of the correspondent landmark points was displayed. In Figure 19 is, then, graphically represented the movement of left and right side markers, while smiling, taking into account their *x*- and *y*-coordinates for the two types of algorithms tested (*Maximal Static Response Assay* (MSRA) and AFA). In addition, the mean difference between these methods was also calculated for each feature, during specific movements, providing a measurement for a particular feature and a mean value for the overall movement. Although simple and fast, this method does not provide a very useful tool for measuring facial motion, in the medical context, in particular for visual analysis through an interesting interface.

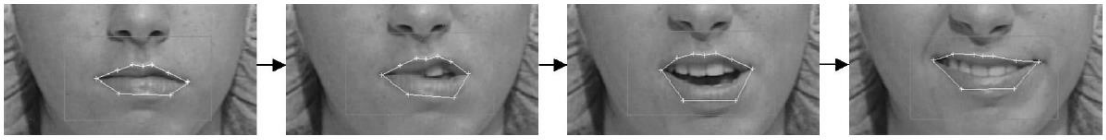


Figure 18. Example of feature point tracking by AFA, in a patient with facial nerve palsy. Retrieved from [21].

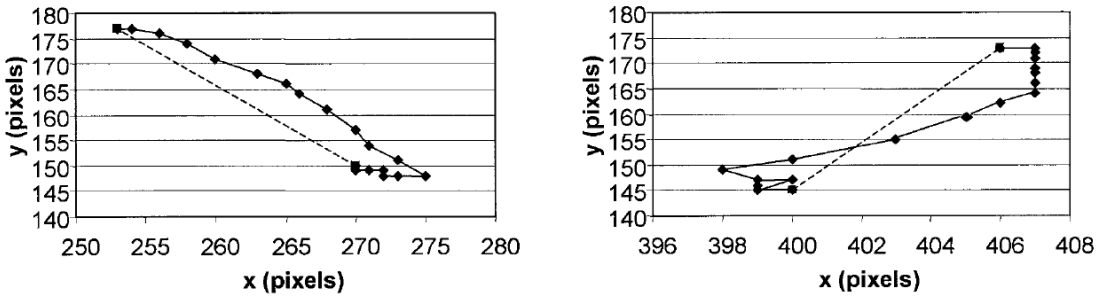


Figure 19. 2D *x* and *y* coordinates for left and right markers in the image sequence shown in Figure 18. Paths of these markers are represented by the lines, using the MSRA method (■) and the AFA method (◆).Retrieved from [21].

Another attempt to create a tool that quantifies the face excursion and, at the same time, providing a visual representation to visualize this movement, was proposed by Tzou et. al. [1]. In addition to the results obtained using the methods described in section 2.4.2, two-dimensional trajectories of each landmark in movement were plotted, through *FaciShow* software. The landmarks' paths were obtained from the feature points' coordinates and, subsequent, the distance between two feature points was calculated. An example of the graphical representations is shown in Figure 20. These can be considered dynamic representations insofar as they seek to represent the motion during smiling, over time, taking into account the distance between two landmarks, in this particular case. As above, this method does not allow a simple and obvious analysis of the facial landmarks trajectories, along the movement and, therefore, should be improved in order to fully characterize this motion and to have medical usefulness.

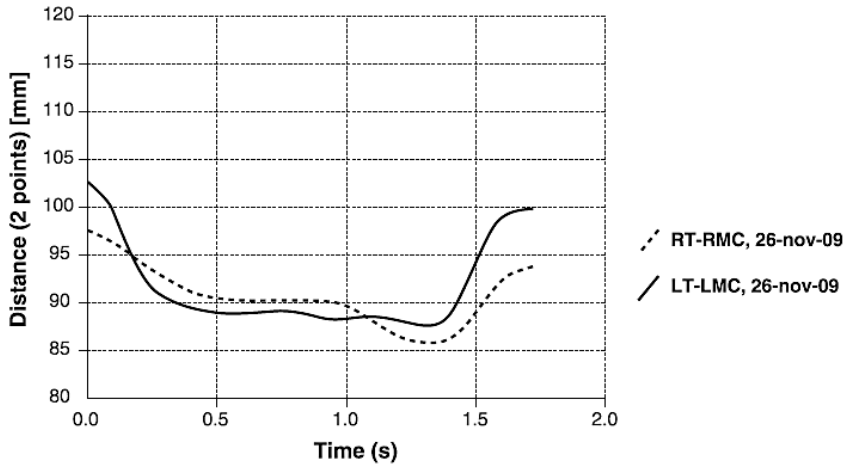


Figure 20. Graphical representation of smile movement, showing the distance/time between landmarks. RT – right tragus; RMC - right mouse corner; LT - left tragus; LMC – left mouth corner. *x*-axis represents time in seconds and *y*-axis is the distance between the two landmarks, in millimetres. Retrieved from [1].

Finally, a more comprehensive approach was achieved by Horta et. al. [5], also described in section 2.5.2.1. In this case, a new method was developed to allow the visualization of dynamics data, using the positions of each landmark previously calculated. These representations contain the paths of each feature point during different movement stages (see Figure 21). In addition, the dynamics of movement can still be represented by horizontal

and vertical displacement of each point as a function of time, giving information about the temporal coordination of each point (Figure 22), and even by differential horizontal or vertical displacement as a function of time, which allows the evaluation of symmetry. Although unique, concerning the display of each landmark path over time, with the possibility of direct measurement of the movement using the graphs' scales, this system is not an adequate tool for, in a simple and obvious way, assess the facial excursion, having a poor interface. Furthermore, this system was designed to make the tracking of the mouth area, during smiling, and needs to be improved to allow the full characterization of the facial movements.

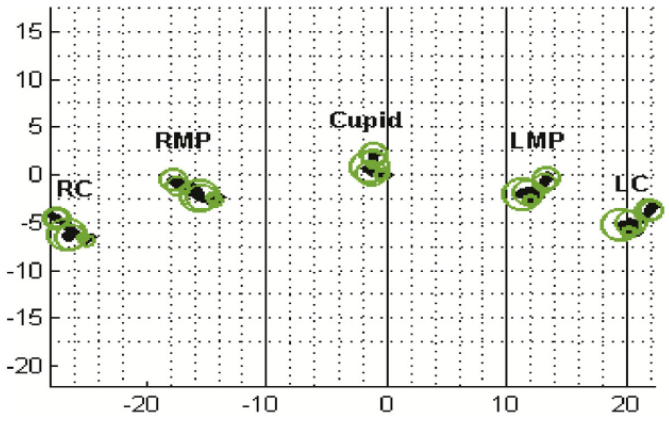


Figure 21. Dynamic representation of the movement of five feature points, for a normal subject. RC - right commissure; LC - left commissure; RMP - right midpoint; LMP - left midpoint. Adapted from [5].

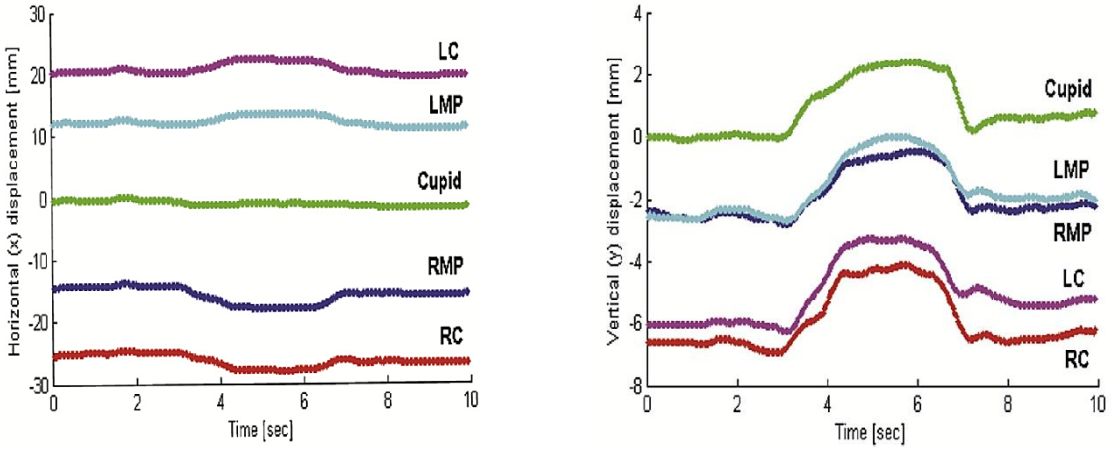


Figure 22. Dynamic representation of a normal subject, showing the horizontal and vertical displacement for five feature points. RC - right commissure; LC - left commissure; RMP - right midpoint; LMP - left midpoint. Retrieved from [5].

2.7.1 Summary

The assessment of normal and pathological behavior associated with facial movements is important so as to analyze the functionality recovering after surgery. Only an objective and quantitative 3D evaluation allows a complete assessment of facial features excursion and it can be achieved by developing a system capable of, simultaneously, capturing 3D spatial-temporal data and processing it in order to facilitate analysis and interpretation. Furthermore, an ideal system should also be simple, easily interpretable, inexpensive, with low computational cost and accurate in order to be sensitive to motion variations with time.

With this in mind, research has been done in order to develop a set of morphological, static and dynamic measurements, also called *facegram*, enabling spatial-temporal and quantitative analysis of facial movements. However, all the presented studies lack simplicity and objectivity in the data presentation, which is a very important feature concerning the medical field. So, this absence of a practical and standard clinical tool in this area leads to the need for a simple and objective system for the assessment of facial movement able to improve existing methods and to kick-start further studies in this context.

Chapter 3

Tracking of Facial Movements

As was detailed in Chapter 2, the goals of the proposed work were the development of a simple and objective system for facial movements' assessment and quantification, intended to be a practical and standard tool. Always bearing in mind the medical barriers and clinical usefulness, several methods were combined to fulfil the proposed objectives and will be explained in detail in the current chapter. First, the processes of data collection and data set creation will be described, so important for the validation of the proposed method. The tools for pre-processing the images' sequences and tracking facial points will also be explained, as well as the approaches to construct facial points' trajectories. Finally, the main results of the described methods as well as the discussion will be considered in this chapter, followed by the main conclusions.

3.1 Data Collection

The evaluation of the algorithms and techniques used in the proposed research depend on the collection of data sets, whether recorded in a clinical environment or not. For this reason, the data acquisition under different conditions and containing different individuals was performed. A depth camera with *Kinect* technology was used to achieve this goal.

3.1.1 *Depth* camera technology

Based on the same sensor present in the *Microsoft Kinect* technology, already explained in Chapter 2, and also developed by *PrimeSense*, the *ASUS Xtion Pro Live* camera[1] (Figure 23) is another motion-sensing device that captures RGB images along with pre-pixel depth information. Being similar to the previous one, this camera generates depth-maps with an operation range from 0.8 to 3.5m and a maximum of 1280x1024 pixels, with 30 frames per second acquisition rate.

[1] http://www.asus.com/pt/Multimedia/Xtion_PRO_LIVE/

Figure 23. *Asus Xtion Pro Live* camera, with appearance and components similar with operation and appearance similar to the *Microsoft Kinect v1* sensor[1].

For this research, an *ASUS Xtion Pro Live* camera was kindly provided to acquire all the necessary data to carry out the project and, lastly, to create a RGB-D-based system for 3D quantification of facial movements.

As mentioned in Chapter 2, there are few or even no researches that incorporate depth measurements for characterization and quantification of facial movements; still, in the present work, the simultaneous acquisition of depth and colour information is suggested as a potential method for the improvement of facial points' tracking, even if the displacement of facial points is around millimetres. For this reason, the depth cameras are a critical resource to carry out the project.

There are already several easy-to-implement and widely used *Natural User Interface* (NUI) libraries and drivers, such as *Microsoft Kinect* SDK and *OpenNI*, which can be utilized to acquire and process RGB-D images, for *Kinect*-based cameras. However, the disadvantage of having a limited language support (mostly, C/C++ and C #) makes them unviable regarding the research scope. Furthermore, taking into account the depth camera that has been made available, not all of these NUI are compatible with the equipment. Therefore, a free available *MATLAB* wrapper[2] that provides an interface to RGB-D cameras through *OpenNI* (v2.2) was used to collect the images from the sensor. In addition to enabling the acquisition of frames, this interface allows the user-modification of settings, such as both depth and color resolution and RGB-depth registration, allowed by the camera's firmware.

In this particular case, after studying the influence of the RGB resolution on camera performance and concluding that the frame rate drops abruptly when using the maximum RGB resolution of the camera (1280x1024 pixels), the reduced 640x480 pixels resolution was used throughout all acquisition processes, preserving a frame rate of about 30 frames per second, essential for capturing facial movements.

Finally, the *MATLAB* programming language was chosen for the implementation of all operations, algorithms, methods and interfaces throughout this research, since, particularly for Image Processing purposes, it has advantages over other programing languages: availability of a very large library of built-in algorithms for computer vision applications; ease in prototyping processes, whether for graphics and GUI design, and facility in algorithms testing without previous recompilation. Furthermore, the need to develop a medical representation of the results (*facegram*) leads to the necessity of non-trivial graphics capabilities, which are readily available in *MATLAB*.

[2] http://www.mathworks.com/matlabcentral/fileexchange/42127-matlab-wrapper-for-openni-2-2

3.1.2 Application for medical data acquisition

In order to facilitate data acquisition in a medical environment, an application was developed and its GUI is shown in Figure 24.

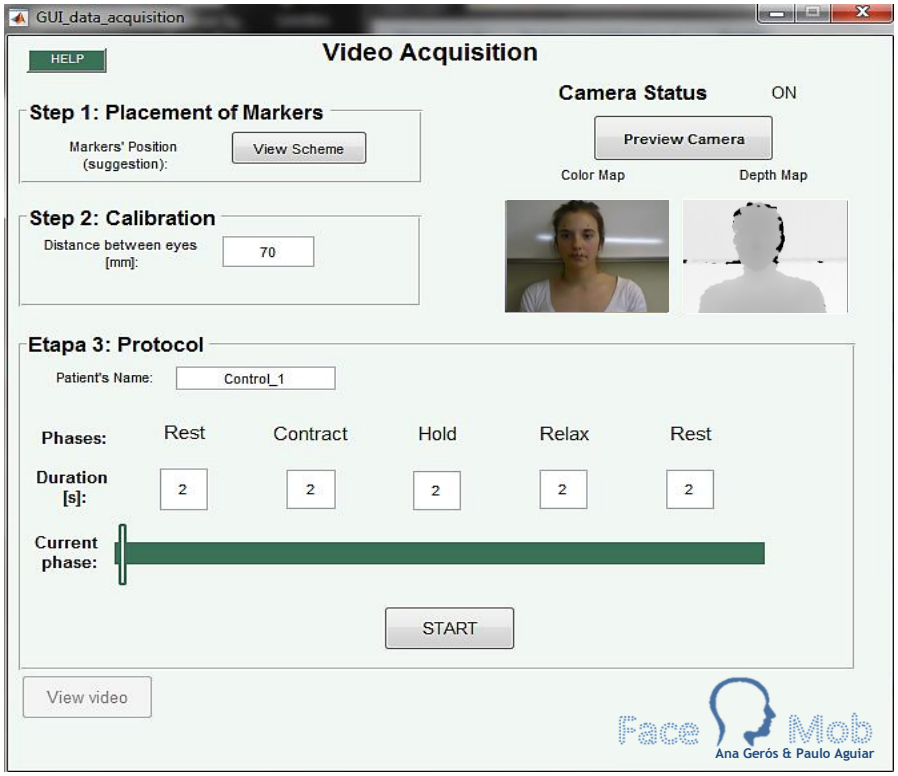


Figure 24. *Graphical User Interface* of the application developed for medical data acquisition (RGB-D data).

This application allows the acquisition of RGB and depth frames, at a maximum frame rate of 30fps, and record the data to a file that will later be used to track the facial points, construct trajectories and characterize the movement. Additionally, the application allows the user to introduce the patient's name and the distance between the eyes (or the distance between two reference ink markers) that will be used as a scale factor control in the conversion of pixels to appropriate measurement units. For each face movement's phase, the user must enter the desired time, to be used as reference for the duration of each movement, when recording. Moreover, the current color and depth map, captured by the camera, can be visualized, as well as a support tool for the placement of the markers on patients' face.

3.1.3 Datasets

Since the present work is truly dependent on the availability of medical samples, essential to clinically validate the results, RGB and depth images of 2 patients were acquired and recorded in medical environment, while the patient was performing specific face movements (see examples in Figure 26). The clinical situation of each patient can be found in Table II. In fact, the acquisition process was possible thanks to the collaboration of Dr. Ricardo Horta, a physician of the *Plastic Surgery, Reconstructive and Maxillofacial* center (SCPRMF), at *Hospital São João*, Porto, and his patients.

Table II. Clinical situation of each subject.

Patients #	*Pathology*	*Sex*	*Age*	*Aetiology*	*Side*	*Surgery performed*
1	Yes	Female	49	Congenital (birth trauma)	Left	Facial reanimation with temporal flap and *fascia lata* graphs
2	Yes	Female	44	Middle ear infection (*otitis media*)	Right	Facial reanimation with *gracilis* muscle transplantation and motor nerve coaptation to the *masseteric* branch
Control 1	No	Female	22	_____	____	_____
Control 2	No	Male	38	_____	____	_____

This study will focus on assessment of smile, in resting and dynamic conditions. as mentioned in Chapter 2, symmetry at rest and the ability to produce a symmetric smile is of key importance in facial reanimation procedures; however, other points' constellations can be studied with the proposed methods described later (the markers can be placed in any position in the face that the clinician finds of interest). For this reason, in the video recording sessions, small ink dots were marked on 5 specific anatomical landmarks: 2 commissures (left – LC and right – RC), 2 midpoints (left – LMP and right RMP) and cupid, and 2 reference markers. (Figure 25).

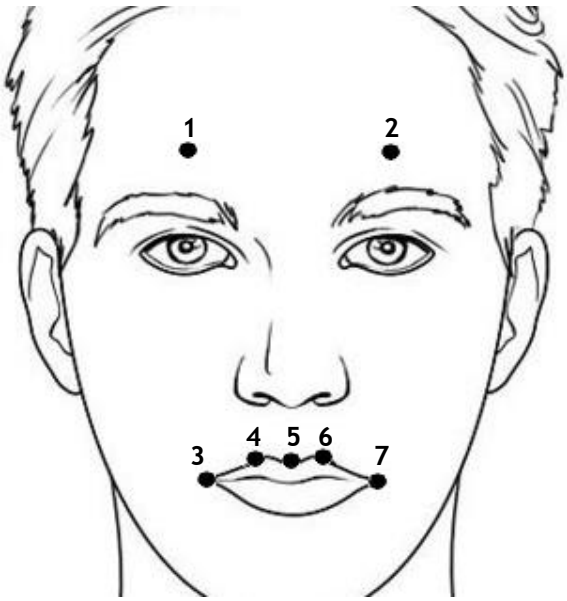


Figure 25. Example of anatomical landmark constellation used in the acquisition of RGB and depth images. It is composed by two reference points (points 1 and 2 – Ref. 1 and Ref. 2, respectively), two commissures (points 3 and 7), two midpoints (points 4 and 6) and cupid (point 5).

As previously referred, ink dots were chosen, as opposed to physical markers, in order to not inhibit the natural facial movements. The use of markers is important from the medical point of view, to facilitate the visual tracking of a particular point at a specific and precise location. Besides that, other reason of requiring markers is taken from the fact that in many facial reanimation situations, the assessment needs to be performed in locations which are not associated with standard facial landmarks. Freedom in the choice of facial points to track

is vital (the physician may want, for example, to assess a muscle group on the check, which is devoid of clear landmarks).

During the recordings, the patient was asked to smile, avoiding head movements, in a series of 5 phases with durations defined by the physician/user: *Rest* (3s), *Contract* (3s), *Hold* (3s), *Relaxation* (3s) and *Rest* (3s) (see Figure 26 for some examples). It should be noted that the patient was positioned with the face centered in camera's plane and with camera's optical axis perpendicular to the face plane. In addition, lighting conditions and color markers were chosen ensuring maximum contrast between the markers and the patient's face.

In addition to data from patients with pathological conditions, additional RGB and depth images were acquired from 2 subjects without known pathologies (controls) for intermediate validation and theoretical trajectories plotting purposes. The clinical situation of these two subjects can also be found in Table II. The acquisition conditions previously described were maintained and the duration of each movement phase was as follows: *Rest* (2s), *Contract* (2s), *Hold* (2s), *Relaxation* (2s) and *Rest* (2s).

All the acquisitions of raw RGB and depth data using the depth camera followed the protocol described in Appendix B.

The best collections of images for each subject (controls and patients) were selected according to the quality of their frames, i.e., the ones that apparently represent larger movements with less light interference and enhanced visual contrast. Figure 26 illustrates examples of frames from each subject (2 patients and 2 controls) with respect to rest and maximum extension phases (one frame from *Rest* and *Hold* phases).

Figure 26. Illustrative examples of acquired RGB images at *Rest* and *Hold* phases, for both patients and controls.

3.2 Methodology

3.2.1 Depth Precision

The accuracy of depth measurements is an important consideration to bear in mind when trying to perform points' tracking using, simultaneously, color and depth information in movements of small amplitude, such as facial movements. Thus, the first experiments started measuring the accuracy of depth measures obtained by the camera. With this is mind, frames from a static target (placed in a rigid support), with three concentric circles (whose diameters were, respectively, 1, 5 and 10 cm), were captured (Figure 27 a)). Knowing the distance from the camera, determined a priori (80cm ± 0.1cm), it was possible to extract, using the obtained 3D information, the depth values inside the circles, supposedly at the same distance. These values were compared with the theoretical one.

In order to calculate the depth measurements' distribution on the plane and time, frames from the same target, described above, were again captured also at a minimum distance of 0.8m from the camera. Thus, the spatial distribution of depth in the three circular coronas was separately determined, trough pixels' values that were within them. In addition, the temporal distribution was calculated by determining the depth value from a set of random pixels along the captured frames as well as the distribution of one particular point through all frames.

3.2.1.1 Scaling factor of the depth camera

For the conversion of the anatomical points' positions, that are in pixel-space, to real-space (in millimeters), it is necessary to obtain a scaling factor. The shorthand to get this factor is to calculate the distance in millimeters, and later in pixels, between two reference markers and obtain the corresponding calibration factor. However, this method is unreliable since it depends on the precision of the measuring instrument that the physician has available, as well as medical accuracy for this measurement. In practice, this measurement will be highly prone to errors. To overcome these limitations, it is possible to calibrate the camera and calculate its scaling factor (an optical factor) as a function of the measured depth.

Therefore, RGB and depth images from the same static target used in the previous experiment, at 80 and 90 cm from the camera, were captured. Knowing the actual distance between each concentric circle, the corresponding distance in pixels obtained by the captured RGB images, the depth values for each point and the distance from the camera (see Figure 27 b)), it is possible to calculate the scaling factor, specific for each case [distances equal to 80 and 90 cm] and compare it with the one that would be obtained by using a simple interpolation of one of the measurements. The camera optical factor is, subsequently, calculated and allows the direct determination of the scaling factor used to convert the obtained results to real-space, depending on the camera distance. This method assumes that the movement of the target facial landmarks is constrained to a volume with a limited depth, where the scaling factor can be assume constant. This condition is later verified.

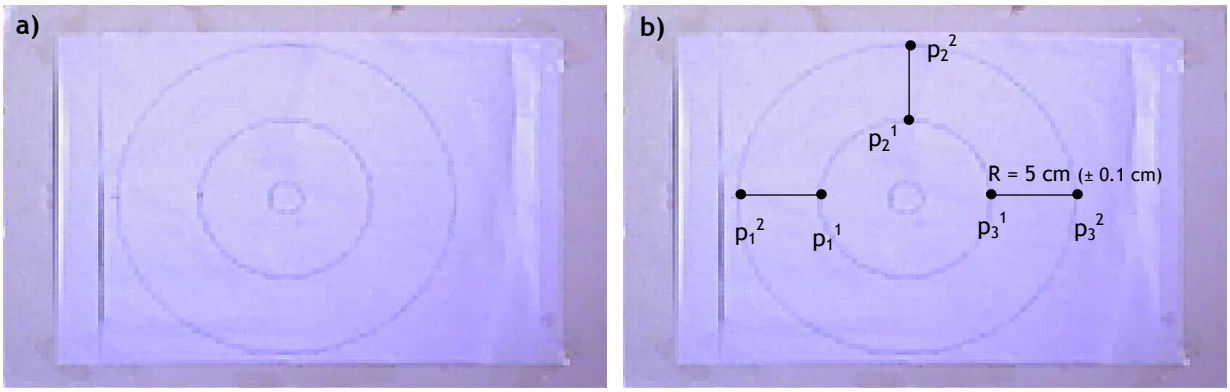


Figure 27. Illustrative example of: a) static target used for depth precision measurements; b) static target with the known distances and necessary measurements in order to calculate the scaling factor. p_x^1 and p_x^2 are the points from two concentric circles with known distance between them, each with 3 samples for statistic comparison.

3.2.2 Depth-Color Registration

The data processing using simultaneously color and depth information is dependent on the registration process, defined as mapping the RGB pixels onto the depth data. Therefore, after modifying the camera settings to allow depth registration, and using both RGB and depth information available, it was possible to determine the alignment vector which registers the RGB and depth frames. For this purpose, depth and RGB frames of an object with well-defined edges were captured and segmented in different ways: the binarization of the depth images was achieved using a threshold defined as the average of the level of background and foreground; the RGB images were converted to binary images using a threshold defined by the user, after defining the region-of-interest (ROI) from user input. Segmentation was, then, the first step to ensure that the alignment vector was calculated without noisy data and from binary images.

In this manner, the intensity-based automatic image registration process could be used to extract the alignment vector that registers the two images. The metric and optimizer parameters need to be specified and can be defined, respectively, as a quantitative measure of similarity between the two images and the methodology for minimize or maximize this similarity metric. Besides that, another information that is required is the transformation type, that defines the 2D transformation capable of align the moving image (in this case, the depth image) with the fixed image (also called, reference image – RGB image). The rigid transformation, consisting of both translation and rotation, was the chosen geometric transformation to be applied to the moving image and, together with a determined transformation matrix, initiates the registration process and determines the transformation that will be applied to the moving image using bilinear interpolation. Using the metric parameter, the moving image, now transformed, was compared to the fixed one, and a metric value was calculated. When the algorithm reaches a stop condition, evaluated by the optimizer, the process ends, resulting the recorded image; otherwise, the process continues to a new iteration and a new transformation matrix is computed. The used algorithm (already implemented in *MATLAB*) is schematized in Figure 28. The geometric transformation, i.e., alignment vector that registers the two images is estimated using the same previous

parameters. This object was then used as a correction factor in subsequent processing, calculations and methods.

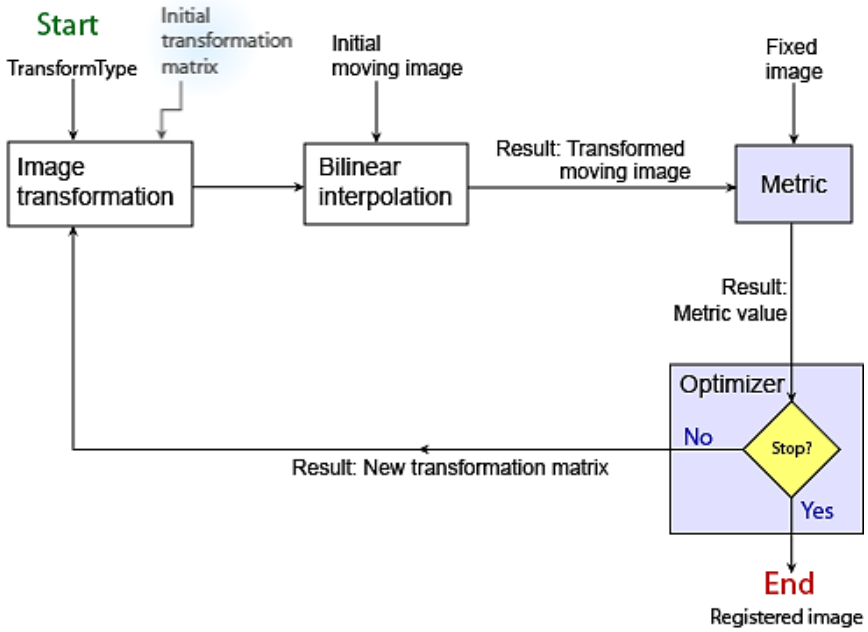


Figure 28. Illustration of the depth-color registration algorithm to obtain the final registered image. Adapted from [103].

3.2.3 Pre-Processing

3.2.3.1 Depth-data recovering

As mentioned, the depth camera sensor provides dense depth estimates (see example in Figure 29); however, the 3D mapping may present noise-related issues, particularly in the depth map, caused by working at a limit distance (if the objects are too far or too close from the sensor), bad angle or surface composition, occlusions, multiple reflections or by objects with particular reflections, such as the human skin or hair, as can be seen in Figure 29. When this happens, the depth map shows regions with unknown depth data (blue pixels) that need to be controlled to prevent errors in subsequent methods. To overcome that, a method for recovering missing depth data was implemented. There are studies in this area with proposed methods to retrieve depth information, such as the work of *Yang et al.* [104]; nevertheless, the method developed in this study allows for a more appropriate, while less complex, approach to address the problem. The method consists on the elimination of zero-depth values inside the frames, replacing them by the depth values of points with the same coordinates but in immediately preceding o subsequent frames.

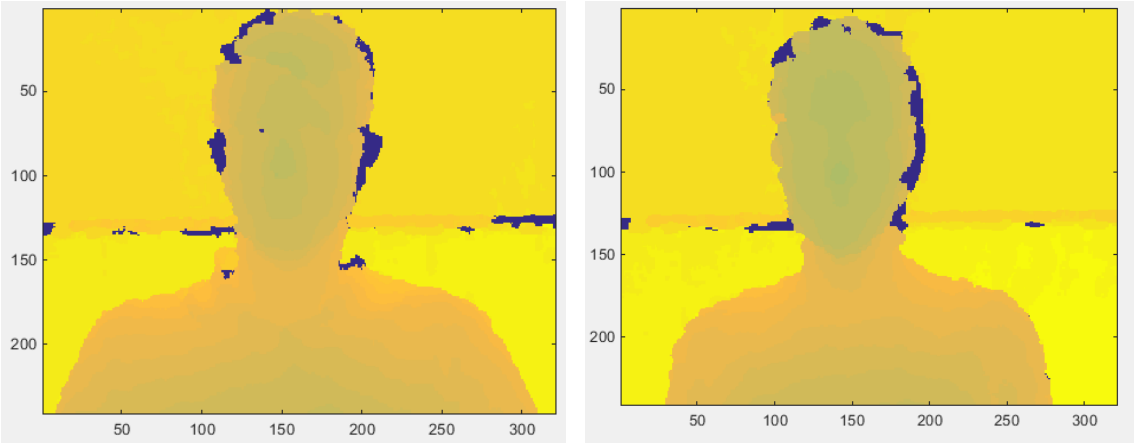


Figure 29. Depth information estimation from 2 subjects, captured with the depth camera. Blue spots in the images represent unknown depth data (cropped images). Distance from the camera: 0.80m.

First, and assuming that the location of the point with missing depth value is (f_1, r, c), being r and c the y and x coordinates, respectively, in frame f_1, points in the remaining frames (f_n) of type (f_n, r, c), with the same point location of the missing point but with non-zero depth values, are selected. Then, to decide between the multiple points, if it is the case, the nearest frame is selected and its depth value is assigned to point in (f_1, r, c).

Unlike the method stated above, this one takes into account the nearest frame information, whose depth value is normally closer to the one that is missing. An example could be an abrupt depression area that contains the missing pixel, whose information cannot be captured by the neighborhood pixels but can be identified in previous or subsequent frames.

3.2.3.2 Color-map contrast enhancement

Once it is intend to use ink-markers pre-placed on the patients face during acquisition, it is necessary to ensure that the contrast between the markers and patient's skin is maximized, so that the tracking is as efficient as possible. Thus, a method to transform RGB images into color-information images was implemented. This method consists in calculating a new image in which each pixel is the distance in color space to the color of the points (markers) chosen by the user (p_{rc}^{m}), in terms of RGB information, where r and c stand, respectively, for the y and x coordinates and m respects to the chosen marker. First, a template is constructed for each pre-selected marker, as illustrated in Figure 30, where T is the half-size of the square template that limits the neighborhood of the selected point (p_{rc}^{m}) and is defined by the user.

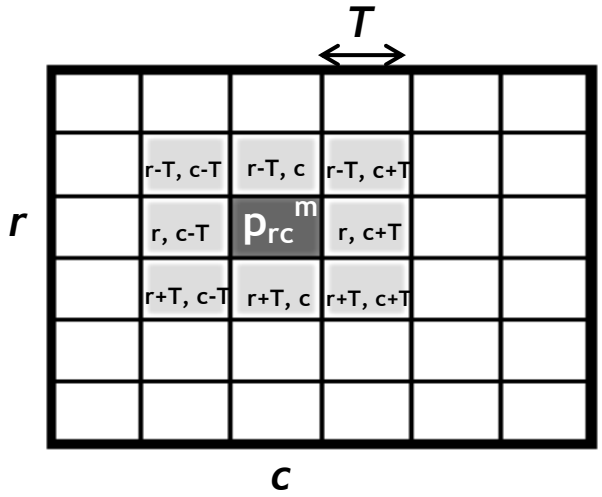


Figure 30. Template around a pre-selected point (marker), used to calculate the median of the neighbourhood.

Then, the median between the neighborhood points of all markers is calculated, for each color channel, separately. This statistical measure was selected to overcome the problem of having too many different intensity values in the neighborhood of the central point. Therefore, reference values are obtained for a specific channel (R_0, G_0, and B_0) and the *Euclidean* distance of these reference points to each image pixel is computed to obtain a new matrix of intensities. These distances can be calculated as follows:

$$d_{RGB}(x_f) = \sqrt{(R_{x_f} - R_0)^2 + (G_{x_f} - G_0)^2 + (B_{x_f} - B_0)^2} \quad (1)$$

where R_{x_f}, G_{x_f} and B_{x_f} are, respectively, the values of red, green and blue channels of the image pixels, R_0, G_0, and B_0 are the reference values of red, green and blue channels, respectively and $d_{RGB}(x)$ is the *Euclidean* distance between these two groups, for a specific point ***x*** at a particular frame ***f***.

In addition to this approach, a different method using the average as a measure to calculate the reference values, instead of median, was tested to verify the performance with respect to tracking efficiency and contrast, as well as only filtering the red channel, without any additional processing. These tracking performance tests will be described later.

For contrast measurement in the images affected by different pre-processing techniques, the quality criteria *Mean Squared Error* (MSE) was used to measure the difference between the mean square errors. In this particular case, the error is the difference between the point of a particular marker (central point) and the pixels around it, with a neighborhood *L*. The images to be compared were first converted to grayscale and, then the MSE was calculated using the following expression:

$$MSE = \frac{1}{L} \sum_{x}^{L-1} \sum_{y}^{L-1} \left(p_{f_1,x_1,y_1} - p_{f_1,x,y}\right)^2 \quad (2)$$

where *L* is the side of the neighborhood around the central point (p_{f_1,x_1,y_1}) in frame 1, and $p_{f_1,x,y}$ are the pixels being compared.

In order to further improve the enhancement of RGB images, a technique, which combines edge detection, using the *Laplacian*, along with a *Gaussian*, to reduce sensitivity to noise before enhancement, was applied to the obtained matrix of intensities. This technique is called *Laplacian of Gaussian* (*LoG*) and its fundamental characteristics are the following: the smooth *Gaussian* filter allows the removal of high frequency noise; the enhancement step is done using the 2D isotropic measure of the 2nd spatial derivative of the image (*Laplacian*); the detection criteria is based on the presence of a zero crossing in the 2nd derivative, combined with a corresponding peak in the 1st derivative; finally, the edge location can be estimate using sub-pixel resolution by interpolation. In the context of the present project, this filter was implemented in order to highlight areas with fast intensity changes (edges), in this case, the markers' boundaries, therefore, enhancing he markers visibility. The *LoG* scale parameter is tuned to the markers typical size (a few mm).

Accordingly, the 2D *LoG* function centered on zero and with *Gaussian* standard deviation σ is defined as follows:

$$LoG\ (x,y) = -\frac{1}{\pi\sigma^4}\left[1-\frac{x^2+y^2}{2\sigma^2}\right]e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (3)$$

The *Gaussian* filter is defined by σ, and the amount of blurring and consequent noise filtering depends on its value. Thus, there must be a trade-off between noise removal (better for larger σ) and edge enhancement (better is smaller σ). In this research, it was used a kernel [8x8] and standard deviation equal to 2 (calculated taking into account the estimated markers' radius). The influence of *LoG* filter on the performance of pre-processing and tracking algorithms was also tested.

3.2.4 Tracking of facial points

To perform tracking, the first step is the construction of two sets of images or blocks required to integrate the tracking algorithm. The first, called reference image or template, is used to search for correspondences in the second set of frames, the search block, i.e. the image to be compared.

Assuming that $p_{f,x,y}{}^{m}$ concerns the new position after tracking the marker *m*, at frame *f*, with *x* and *y* coordinates, the template, specific for each marker, is always centered on the user-selected point in frame 1 ($p_{1,x,y}{}^{m}$), surrounded by *L1* pixels. In turn, the search block in frame *f* comprises the point's position determined by tracking in frame *f-1* ($p_{f-1,x,y}{}^{m}$) for a marker m, with a neighborhood of *L2* points (see Figure 31), taking into account the implementation of a kinematic model, explained in the following section. The size of the search image was reduced so as to be a square block with half-size equal to *L*2, instead of using the entire image for frame *f*, with the aim of reducing the computational time. This is possible assuming that between frames the movements are smooth and with few variations and that facial landmarks do not overlap in time.

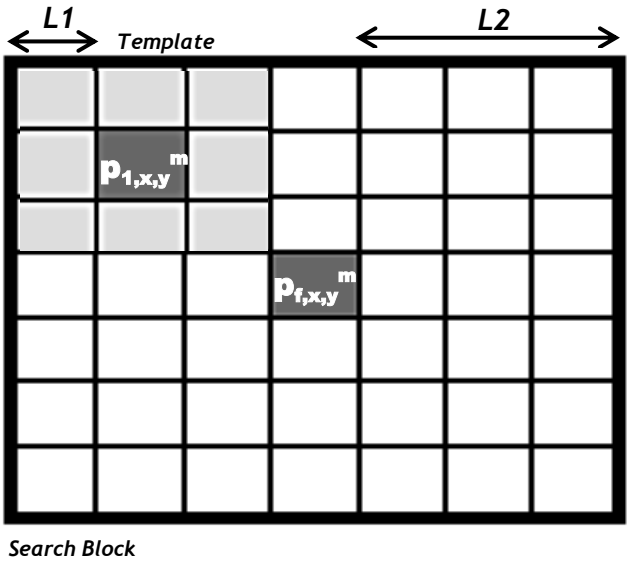


Figure 31. Two sets of fames to be used for tracking, at frame *f*: template and search block.

As mentioned, the template is constructed so as to be always centered on the point in frame 1. However, two alternatives were tested, which consisted on the template construction centered, respectively, on the point of the previous frame or 5 frames ago. For both, the performance of the tracking algorithm was studied.

L1 and *L2* sizes are specified by the user, although with a default value equal to 7 and 11, respectively. The choice of *L1* influences the performance of the tracking and should have the following characteristics: it should not be too small in order to include marker and contextual information, i.e., neighbourhood data (skin information, etc.) and should not be too large so as to include information that does not disturb the tracking or increase the computation time. The size of *L2* should be, for obvious reasons, greater than *L1*.

3.2.4.1 Kinematic model

In order to accelerate the computations, but keeping individual identities, the tracking of multiple points take advantage of internal kinematic models. A simple model allows the determination of the central point in the search block, taking into account the kinetic information of the point in previous frames.

Therefore, an estimate of the point at frame *f+1* is achieved so that the point's position in the previous frame (*f*) is added to the relative position vector $\boldsymbol{v}_{f-1,f}$, according to Equations 4 and 5, and as illustrated in Figure 32. This vector defines the position of a point relative to another one in a different frame, and is the difference between the position of the two points.

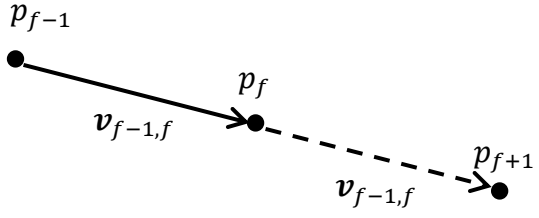


Figure 32. Simple kinematic model for point's estimation at frame *f+1*.

$$\boldsymbol{v}_{f-1,f} = p_f - p_{f-1} \quad (4)$$

$$p_{f+1} = p_f + \boldsymbol{v}_{f-1,f} \quad (5)$$

where p_{f+1} is the estimated point in frame *f+1*, determined by the sum of the point's position (p_f) in frame *f* and the relative position vector $\boldsymbol{v}_{f-1,f}$. It should be noted that this method will take advantage of the fact that facial landmarks do not overlap in time.

Thus, the set of frames that will be compared with the template is centered, for a given marker, on point p_{f+1} with half-size equal to *L2* and will be subsequently used to perform tracking.

3.2.4.2 Block Matching algorithm

The algorithm chosen to perform the tracking of facial points is called *Block Matching* algorithm. As previously explained in Chapter 2, this algorithm is used for motion estimation, the process by which motion vectors are calculated and describe the transformation that occurs from a 2D image to another, in this case, between adjacent frames in a set of images. The blocks intended to be compared are the ones described in the previous section: the template and the search block. In order to compare the matching between the block of pixels in frame *f* to the block in frame *f+1*, cost functions are used, such that the block that results in the least cost is the one that closely matches the current block and, thus, contains the tracked point (see Figure 16). The cost function chosen to be minimized is *Mean Absolute Difference* (MAD), given by the following equation:

$$MAD = \frac{1}{L^2} \sum_{x}^{L-1} \sum_{y}^{L-1} \left| p_{f+1,x,y} - p_{f,x,y} \right| \quad (6)$$

where *L* is the side of the reference block, $p_{f+1,x,y}$ and $p_{f,x,y}$ are the pixels being compared in block of frame *f+1* and reference block in frame *f*, respectively.

The way the reference block searches for the block of pixels in the next frame and, so, finds the best possible match, depends on the search method. In this case, the algorithm calculates the cost function at all possible locations in the search image and the size of the search block is proportional to the number of computations. This method is *called Exhaustive Search* method.

Once the acquired frames have both color and depth information, minimizing the cost function must be done in a 3D environment. In fact, in terms of implementation, the original BMA has been modified to combine both the color and depth information. Thus, the tracking point was done in three dimensions, which is essential to fully characterize the complex facial movements.

Therefore, the MAD for color and depth images was calculated separately and then the sum of both errors was performed. Finally, the minimum of this sum was determined, obtaining the center point's coordinates of the matched block that minimizes it. In cases where two pairs of points' coordinates in frame *f* are obtained and minimize the cost function, the implemented algorithm selects the one that minimizes the distance between the coordinates of this frame *f* with those obtained for the frame *f-1*, assuming again little variation between frames.

In addition to this minimizing method, other 3 were implemented and tested for performance. The first method consists in determining the median around a point in the matrix of sums and the matched point for this frame is the one that minimizes the corresponding median. The other two methods are based on the convolution of a *Gaussian* kernel and the RGB and depth matrices of errors and subsequent minimum calculation. The low pass filter is characterized by a filter size and standard deviation equal to [3x3] and 0.5 (default values), respectively. What distinguishes both methods is that, in the first one, the matrix that is intended to be used for convolving is the errors' sum matrix, and in the other one, the convolution is made using the errors in RGB and depth separately and, then, the sum and minimum are calculated.

3.2.5 Head Movement Correction

In data acquisition, an aspect that has to be taken into account is to ensure the immobilization of the patient's head during facial movements (in this case, the smiling), such that pixels conversion into measurements units is accurately performed. Unfortunately, when acquiring the images in medical environment, this requirement cannot be completely controlled by physicians and repeatedly fails due to patient's conditions. There are strategies that could have been done to keep the patient's head in a stable position (fixing it against a flat surface, for example) but it should be notice that these mechanisms that would guarantee head immobilization would potentially affect the facial movement and render them un-natural.

Therefore, an algorithm was developed to correct the movement of the head and was implemented using the tracked coordinates of the markers in all frames, after tracking. That's because, if the movement of the head is not accounted for, the trajectories of the points are a combination of movement in the face reference frame together with the movement of the head. For this reason, when acquiring the RGB and depth images, beyond the markers placed in the desired anatomical positions, it is necessary to add two other reference markers, placed, for example, in the patient's forehead aligned with the pupils and parallel to each other: when studying the smile contraction/relaxation, the forehead is not subject to muscle movements and the positon of the reference markers can be considered fixed in the head references. In different studies, other locations can be chosen. These reference points are used to compensate head movements in the image plane, assuming rigid body transformations between frames.

The first step is then to obtain the transformation matrix (rotation and translation) characteristic of each frame. This is possible comparing the matrix that characterizes the two reference points in each frame and the matrix of a fixed set of reference previously defined. Assuming that the fixed reference vector is given by:

$$\hat{r} = \binom{1}{0}, \tag{7}$$

a horizontal unit vector that defines the reference direction, and the reference midpoint is the midpoint between the two reference markers in frame 1 (PM_1), the matrix for the fixed reference points is given by:

$$A = \begin{bmatrix} PM_1 \\ \hat{r} + \quad PM_1 \end{bmatrix} \tag{8}$$

where $\hat{r} + PM_1$ respects to the resulting point of the sum of the unit vector with the midpoint of the reference markers in the first frame.

In turn, the matrix specific for the two reference markers in a frame *f* is as follows:

$$B = \begin{bmatrix} PM_{\mathrm{f}} \\ \hat{u} + \quad PM_{\mathrm{f}} \end{bmatrix} \tag{9}$$

where PM_{f} is the midpoint between the reference markers that appear in the images of frame f, $\hat{u}$ is the normalized vector between the coordinates of these markers and $\hat{u} + PM_{\mathrm{f}}$ is the resulting point by summing the PM_{f} with the unit vector. The matrix A is always the same to be used for comparison with B, which varies from frame to frame. These two matrices A and B are compared in order to obtain the transformation specific for each frame.

The transformation matrix is efficiently calculated in MATLAB taking advantage of the properties of the *Singular Value Decomposition* (SVD) applied on the 3x3 matrix defined by equation 10 [105]:

$$H = \sum_{m=1}^{2} p_{m,x,y}\ p'_{m,x,y} \tag{10}$$

where m is the reference marker (1 and 2) and $p_{m,x,y}$ and $p'_{m,x,y}$ are the two corresponding point sets, before and after transformation. These point sets are related by:

$$p'_{m,x,y} = R\, p_{m,x,y} + t + \varepsilon_i, \tag{11}$$

being *R* the standard 3x3 rotation matrix, *t* the 3D translation vector and ε_i a noise vector.

Once obtained the rotation (R_f) and translation (t_f) matrices (see Equations 13 and 14), for each frame f, the corrected positions $p'_{f,x,y}$ for the other markers (placed in anatomical positions) are calculated as follows:

$$p'_{f,x,y} = (p_{f,x,y} - t_f) \times R_t \tag{12}$$

where $p_{f,x,y}$ is the original position of a marker calculated after tracking, in frame f.

$$R_f = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \tag{13}$$

where θ stands for the rotation angle.

$$t_f = \begin{bmatrix} x \\ y \end{bmatrix} \tag{14}$$

3.2.6 Trajectories of facial points

3.2.6.1 Construction of facial points trajectories

After determining the positions of all anatomical markers for each frame, it was possible to build distinct graphic traces, in order to analyze different aspects of facial dynamics.

With the raw data collected, two kinds of traces were constructed, containing different information regarding the paths followed by each anatomical marker.

The first (type 1) is the plot, in scale, of the horizontal (*x*) and vertical (*y*) displacement, i.e., the distance from the starting point (in pixels), of each marker, as a function of time. Such tracing provides dynamic information, for example, the velocity in each temporal section, as well as temporal coordination of each marker and also between the two components [horizontal and vertical], which can be useful in certain types of paralysis where such coordination is affected [5]. These plots can be supplemented with information about depth and, therefore, the trajectories of each marker regarding the corresponding depth values as function of time were also constructed.

The other one (type 2) concerns the parametric paths traveled by each marker, and information such as symmetry in the contraction and relaxation paths, for example, can be extracted from these plots. In each trajectory 6 temporal points were marked: start of recording (T0), start of contraction (T1), end of contraction (T2), start of relaxation (T3), end of relaxation (T4) and end of recording (T5). The parametric paths relate to traces of (*x(t), y(t)*) as a function of *t*, i.e. the relative position or displacement of *x* and *y* coordinates as a function of time, respectively. In order to obtain a continuous path for each marker and taking into account that the obtained data from each position over time is discretized, it is necessary to perform interpolation for each function *x (t)* and *y (t)* and combine the resulting curves to obtain the desired path. The method used to implement the interpolation algorithm will be described in the following section.

3.2.6.2 Interpolation

Interpolation of marker pixel positions was performed with two purposes: i) filter out local fluctuations due to data acquisition noise; and ii) achieve subpixel measurements by transforming the discretized marker pixel positions to continuous values.

The interpolation method chosen to fitting the discretized trajectories of markers over time was the spline interpolation. This is a form of interpolation that uses low-degree polynomials in each of the intervals, resulting in a spline function, i.e., a numeric function that is piecewise-defined by polynomial functions in the aim of maximizing the fitting between the data for each interval. The interpolation using a smoothing spline is typically developed when data is noisy, which is the case of the obtained data in the present research.

Mathematically speaking, the smoothing spline *s* minimizes the following expression:

$$p \sum_i w_i \ (x_i - s(t_i))^2 + (1-p) \int |D^2 s(t)|^2 \, dt \qquad (15)$$

Here, t and x stand for time and *x*- or *y*-position data obtained using tracking, i is the entries of t, $D^2 s$ denotes the second derivative of function *s,* and p and w_i the specified smoothing parameter and weights, respectively, used for the construction of the spline. The smoothing parameter determines whether *s* is smooth or close to the data and varies from 0 to 1. For p = 0, the spline behaves as a least-square straight-line fitted to the data; in turn, when p=1, the interpolation produces a cubic spline interpolant function. In this work, the p value was equal to 0.99, so that the interpolation follows the data points, in order to reveal possible abrupt changes, as is the case of facial spasms, but smooth enough not to be forced to go through all data.

After obtaining a new set of data using interpolation for both the *x*- and *y*-coordinates (*x(t)* and *y(t)*), the two interpolated curves were combined in order to construct the trajectory paths (type 2), described in the previous section.

3.2.6.3 Application for tracking and visualization of medical data

The algorithms described in the previous sections were embedded in an application in order to be used in medical environment and to facilitate data processing and reading, during and after tracking. The GUI of this application is shown in Figure 33.

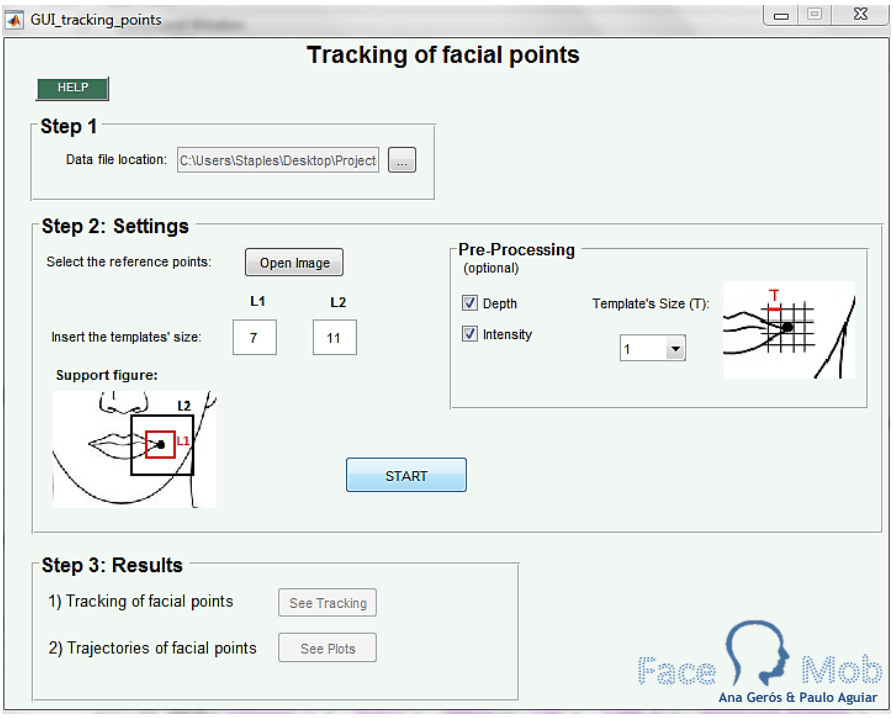


Figure 33. *Graphical User Interface* of the application developed for point's tracking.

Firstly, this application allows selecting a data file to be analyzed in order to perform tracking using a recorded data file, e.g., acquired by the data acquisition application, previously mentioned. Thereafter, the user must manually choose the location of the markers by selecting them over the subject's face projected in 2D image of frame 1, as well as trace the face axis. In order to do that, the GUI provides tools such as pointers, zooming and line axis tracings, to facilitate the user interaction. It is worth mentioning that the user is notified that the facial vertical axis is usually defined by the line that connects the *menton* to the *mid-glabellar* area. Then, the size *L1* and *L2*, which are the half-size of the template and search blocks, respectively, the pre-processing steps (optional) and the size of *T* (half-size of the square template for color pre-processing, if it is the case) are chosen and introduced by the user. After performing the tracking, the user can watch the video that contains the multiple RGB acquired frames as well as the location the tracked points, identified by a red circle, along the frames. Also, both graphs trajectory of points (type 1 and type 2) can be displayed and later saved. All data required for further analysis as well as the video can also be stored.

3.2.7 Validation of the algorithms

As it has been mentioned along the methodology of this research, all the algorithms, and other alternative methods, were tested for performance regarding the tracking of the facial points.

In order to validate these same techniques and select the most appropriate methods, theoretical paths were built through manual follow-up of the points. For this purpose, several sets of RGB and depth images were acquired from two subjects with no apparent pathology, used as control group, as described in section 3.1.3. The best sets were selected (2 for each control), taking into consideration the collection of those in which the subjects performed larger movements. In the case of control 2 (see Table II), in addition to facial movements, the subject moved, excessively, the head during acquisition in order to screen any errors and validate the correction of head movement method.

Thereafter, the coordinates of all markers over all frames, for both sets, were manually determined and theoretical, or gold standard, trajectories were built, for each set of images separately. To be able to compare the real (with computationally tracked points) and theoretical trajectories, the facial points' tracking was done for each set of images, using the algorithms to be evaluated. This validation was done taking into account only the spatial components x and y. Both [real and theoretical] trajectories were compared using two different statistical measures: the coefficient of determination and the mean error. The first one, denoted R^2 is a statistic that indicates of how well data fit a model, a regression line or simply a curve. In this case, R^2 allowed to determine how close the theoretical and actual paths were and the better the theoretical trajectory fits the data, the closer the value of R^2 is to 1.

Assuming a data set with *n* values where t_n corresponds to the theoretical data, each associated with a predicted value (or real value) f_n, it is possible to calculate the R^2 regarding each data set using the following expression:

$$R^2 = 1 - \frac{SS_r}{SS_t} \tag{16}$$

Here, SS_r and SS_t stand for the residual sum of squares and the total sum of squares, respectively, whose mathematical expression are as follows:

$$SS_r = \sum_i (t_i - f_i)^2 \tag{17}$$

$$SS_t = \sum_i (t_i - \bar{t})^2 \tag{18}$$

where $\bar{t_i}$ respects to the mean of the observed data:

$$\bar{t} = \frac{1}{n} \sum_{i=1}^{n} t_i$$

In turn, the mean error between the two data sets (theoretical and real) was also calculated, and it is defined as the average number of pixel that the real data (containing the tracked points) are deviated from the theoretical ones (see Equation 19).

$$meanError = \frac{1}{n} \sum_i |(\ t_i - f_i)| \quad (19)$$

Both statistical measures were separately calculated for ***x*** and ***y*** coordinates and analyzed, together with the graphical representation of both trajectories.

As mentioned, this validation technique was used to compare different methods, which were:

- The average as a measure to calculate the reference value for color pre-processing, or only filtering the red channel as only processing, instead of calculating the median of the central points' neighborhood of the central point.
- The position of each point in the previous frame or 5 frames ago to construct the template block for matching, rather than using the template always centered at the point of the first frame;
- Usage of different methods to minimize the cost function in the tracking algorithm, and
- Specification of different sizes of *L1* and *L2* templates to verify their relevance in tracking.

It is important to notice that the validation of these methods can only be made without the implementation of head movement correction algorithm, since theoretical positions of facial points are determined visual and manually.

3.2.7.1 Validation in a controlled environment

The need to provide a complete validation of the tracking system in 3D space has led to the use of a computer controlled system capable of moving a marker in the three dimensions environment. This system, illustrated in Figure 34 a), consists on a robotic arm working in Cartesian coordinates, to which is connected a sheet with a reference-circle mark, whose coordinates can be controlled by software, in order to follow a well-defined path. The trajectory of the marker was coded using *G-code*, a well-known instructions language used in computer-aided manufacturing for controlling automated machine tools such as industrial robots and 3D-printers. The used system allowed a precision of 100 μm in all three orthogonal directions. The computer-controlled movement of the marker in 3D space was captured by the depth camera and processed, as if it was a facial landmark.

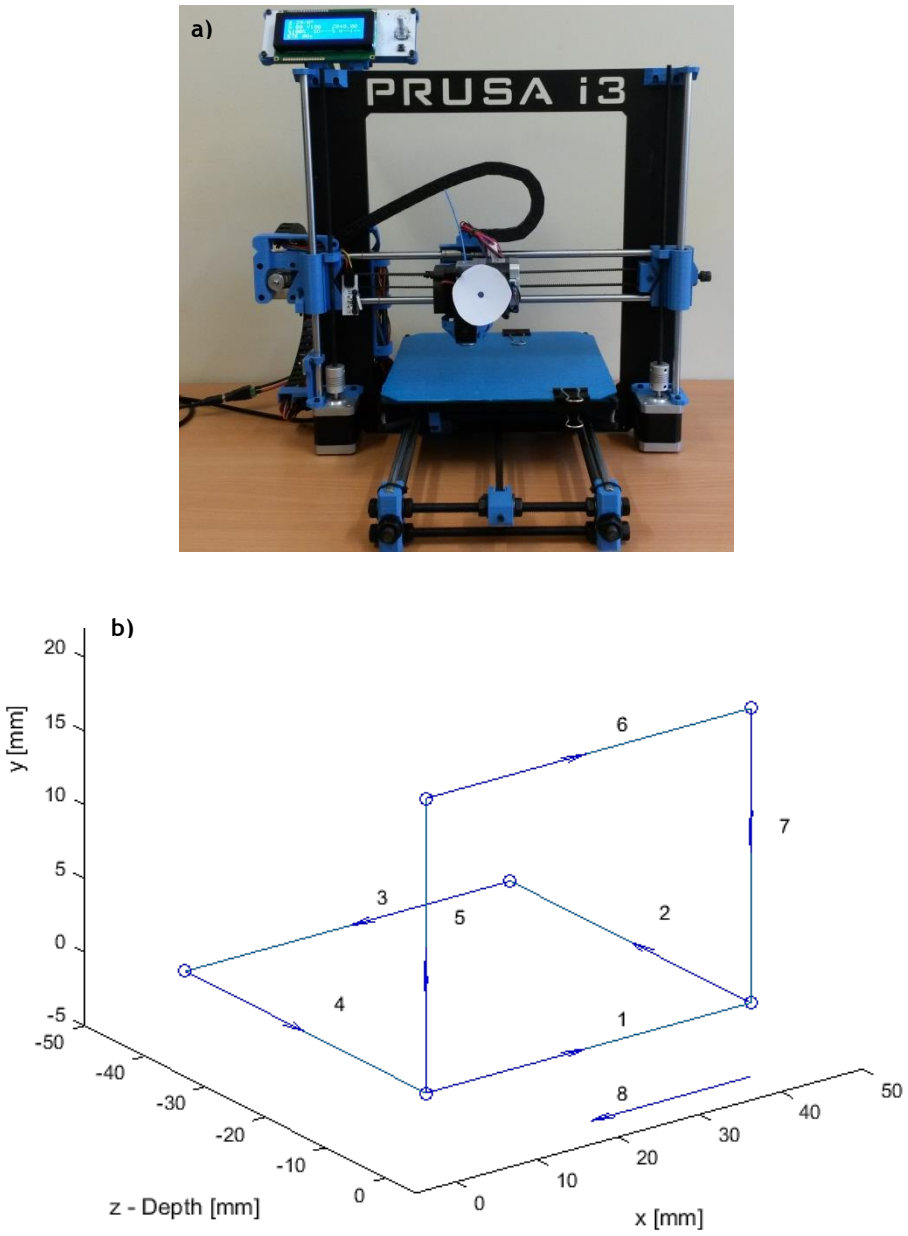


Figure 34. Computer-controlled system, composed by a robotic arm and a sheet with a reference-circle mark (a)). Theoretical path traveled by the robotic arm over the numbered segments (b)). Blue circles respect to the vertices of each segment, pre-defined as coordinates. Distances converted to millimeters.

Knowing the trajectory in millimeters that was programmed with a precision of 100µm (see Figure 34 b)), and the automatically tracked points, it was possible to superimpose both informations in a plot. In the case of horizontal and vertical components (*x* and *y* coordinates, respectively) a scaling factor was applied, which was calculated taking into consideration the distance traveled by the robotic arm on the first section of the path and the corresponding distance in pixels, and assuming that the calibration of *x* and *y* coordinates is equal (standard calibration method). The depth values were corrected to the millimeter scale (data divided by 10; depth in the RGB-D camera used is provided in units of 100 µm). Both theoretical trajectory and real tracked points were once again compared taking into account the R^2 and mean error measures.

Once the data contain *x*, *y* and *z* components, R^2 had to be calculated on 3D environment, also using Equation 16. Firstly, the distance *d* between each experimental point, given by $p_{x,y,z}$, and a straight line defined by the two vertices of each theoretical segment, $t1_{x,y,z}$ and $t2_{x,y,z}$ was calculated using the following expression (see Figure 35):

$$d = \frac{\left|\left(t2_{x,y,z} - t1_{x,y,z}\right) \times \left(t1_{x,y,z} - p_{x,y,z}\right)\right|}{\left|t2_{x,y,z} - t1_{x,y,z}\right|} \tag{20}$$

where x denotes the cross product between the two vectors $\left(t2_{x,y,z} - t1_{x,y,z}\right)$ and $\left(t1_{x,y,z} - p_{x,y,z}\right)$. The SS_r, in Equation 16, represents the sum of these distances *d*.

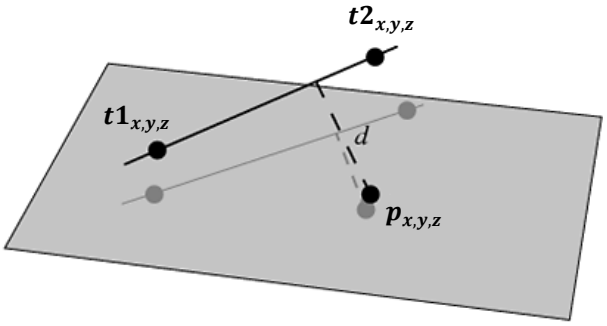


Figure 35. Schematic illustration of the distance between a point and a line in 3D.

The total sum of squares, SS_t, was calculated by taking the square of differences between each tracked point and the average of the set of tracked points. Since in each trajectory' segment only one component (vertical, horizontal or depth) is changed, the mean of the observed data and SS_t were calculated for the component that modifies in each section.

Finally, the R^2 between the theoretical trajectory and real points was calculated for each section separately using Equation 16.

Similarly, the mean error, in this case a 3D statistical measure, was calculated between the line defined by the two vertices of each theoretical segment and each experimental point. The sum of the differences represented in Equation 19 is given by SS_r, already obtained for R^2. Finally, this sum is divided by the number of elements to obtain the average deviation between real and theoretical data, in mm.

3.3 Results and Discussion

3.3.1 Depth Precision

The accuracy of depth measurements was evaluated, by determining the pixels, and their depth values, within the three concentric circles, with radii equal to 1 (R1), 5 (R2) and 10 cm (R3), respectively, and their circular coronas, as shown in Figure 36 a), b) and c).

The mean and corresponding standard deviation of the points inside each circular corona were equal to 81.50 ± 0.14 cm (R1), 81.55 ± 0.39 cm (R1 and R2) and 81.61 ± 0.84 cm (R2 and R3), with depth precision of 100μm. In fact, the distance between the camera and the target was known, and approximately equal to 80cm. This value is impossible to accurately measure since the real distance between depth sensor lens and the target is unknown. Comparing the

real distance and the one obtained from the depth values of points inside the circles, it is observed that the difference is approximately 1 cm. Notice however that absolute depth values are not required in the developed system, only relative depths, so the unknown offset in the depth axis is not relevant.

The important results are: i) the error is the order of 1 mm in the mean depth, between inner circle and outer corona, which is in accordance with what was expected; and ii) the fast signal degradation in terms of noise/fluctuations as one goes way from the optical axis of the camera.

Regarding the distribution of depth data on a plane, and taking into consideration the obtained results previously mentioned, these were organized into histograms (Figure XX13 d), e) and f) to determine the variation of depth data for points that, in theory, were at the same distance from the camera. The differences between the means of points' distribution within the circles increases with the number of points, being approximately equal to 0.6mm between consecutives coronas. Theoretically, the points belonging to the 3 coronas should have approximately the same depth value. In fact, the standard deviation in each circle does not exceed 1cm, being higher for larger radius, and, therefore, the differences are small, possibly negligible. Since the distribution's means are within the error bars, it is not necessary to make depth adjustments (because there is an insignificant lens aberration or the errors are so large that the correction is unnecessary).

The distribution of depth measurements over time was also evaluated and a random set of 11 pixels was selected and their depth values analyzed over 5 and 100 frames, through distribution histograms (Figure 37). The results were similar to the previous, with standard deviations less than 1 cm, showing consistency of depth measures over time.

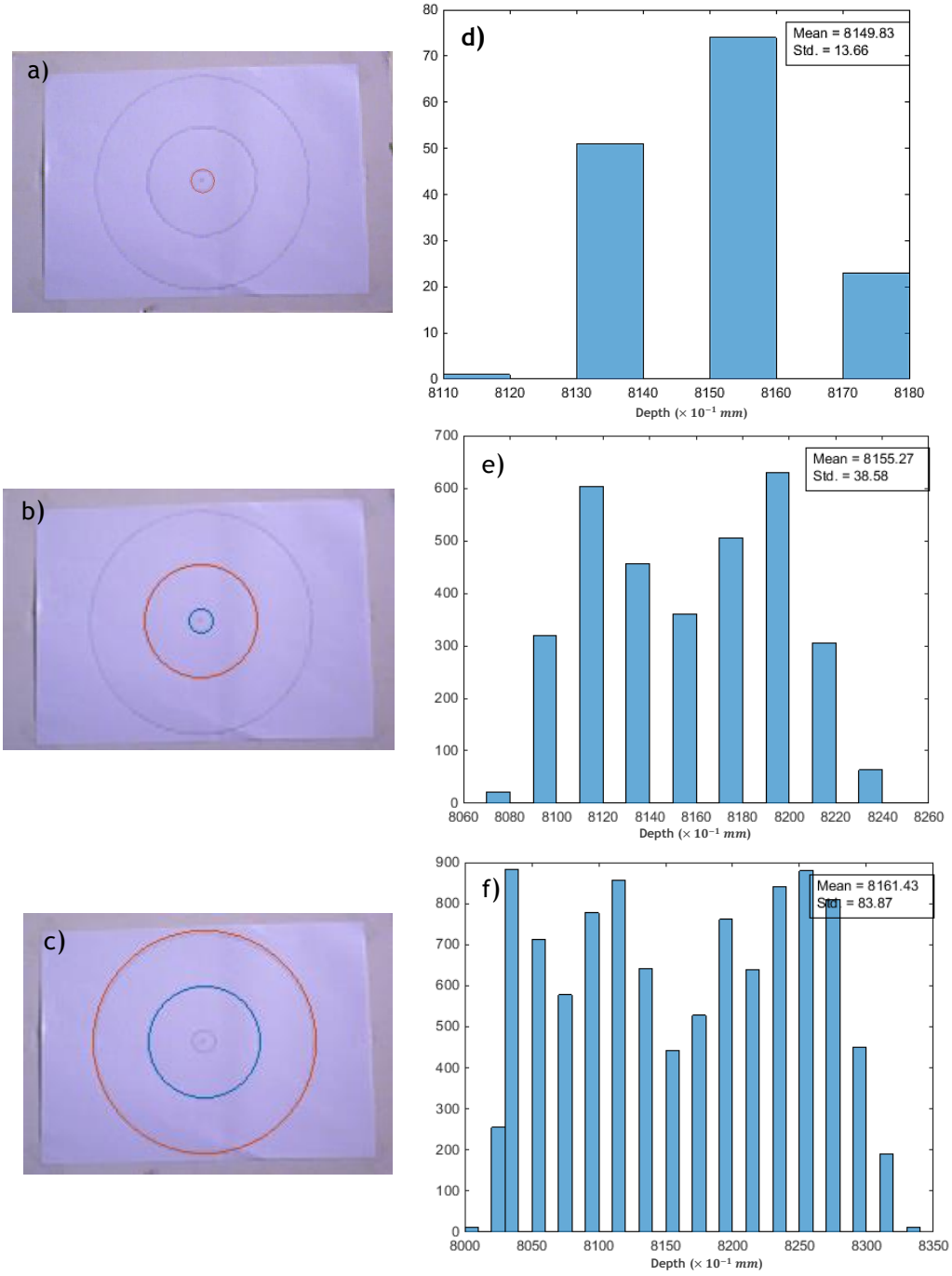


Figure 36. Analysis of depth values' distribution in the 3 circular coronas, separately. a), b) and c) respect to the RGB frames with the 3 circular coronas programmatically identified. d), e) and f) are the histograms of the points' distribution in each circle.

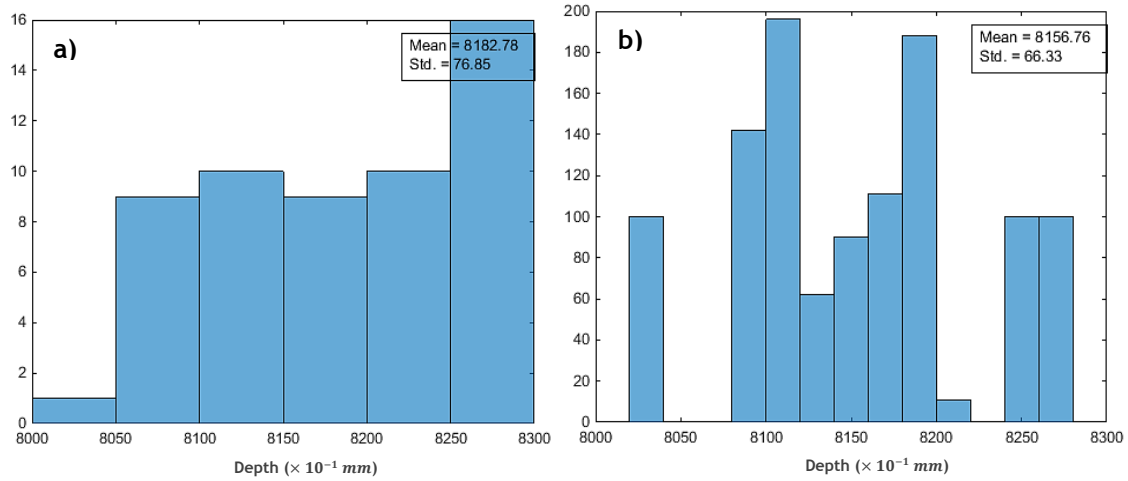


Figure 37. Analysis of depth values' distribution, for a set of 11 pixels, along 5 (a) and 100 frames (b).

Finally, the depth distribution for a particular pixel was tested, along the depth of 5 and 100 frames. The results showed that the mean differences between the two sets were approximately equal to 2mm and the standard deviation less than 1.1 mm. Therefore, a stable behavior of pixels over the frames for a static object was confirmed and depth corrections are not necessary in pursuit of the research objectives.

3.3.1.1 Scaling factor of the depth camera

With the aim of calculating the intrinsic scaling factor of the camera, which will convert the obtained results to real-space, several measures were determined and are described in Table III.

Table III. Required data for calculating the intrinsic scaling factor of the camera.

		Circle points' measurements			
		Pixels	**Distance - R** [mm]	**Scaling factor (theoretical)** [px/mm]	**Depth** [mm]
Distance - camera [mm]	800	32	50	1.56	817.65 ± 3.37
	900	28	50	1.79	915.35 ± 4.76

Assuming proportionality between measurements, and using values obtained for the distance of 80 cm, the specific scaling factor calculated for 90 cm was equal to 1.75, with a relative error of 2.23%. Given the low error found in this distance range, the intrinsic scaling factor of the camera was calculated (0.0019/mm), using only one of the measures. As a result, the camera was calibrated so that for a given distance from the camera, known a priori, the scaling factor used to convert the measures in real-space is easily known, without the need for additional measurements.

3.3.2 Depth-Color Registration

As explained in the methodology section, the wrapper used for data acquisition allows configuring some firmware parameters of the camera, including the activation, or not, of the

depth-color registration. Therefore, after concluding that the correction made by the activation of this parameter did not allow the correct alignment of the RGB and depth images, although slight differences between depth images with, or without, the activation of this command were visible, the alignment vector was calculated using the method previously described. The results from the segmentation of depth and RGB images, captured from a static object, are shown in Figure 38 a) and b), respectively.

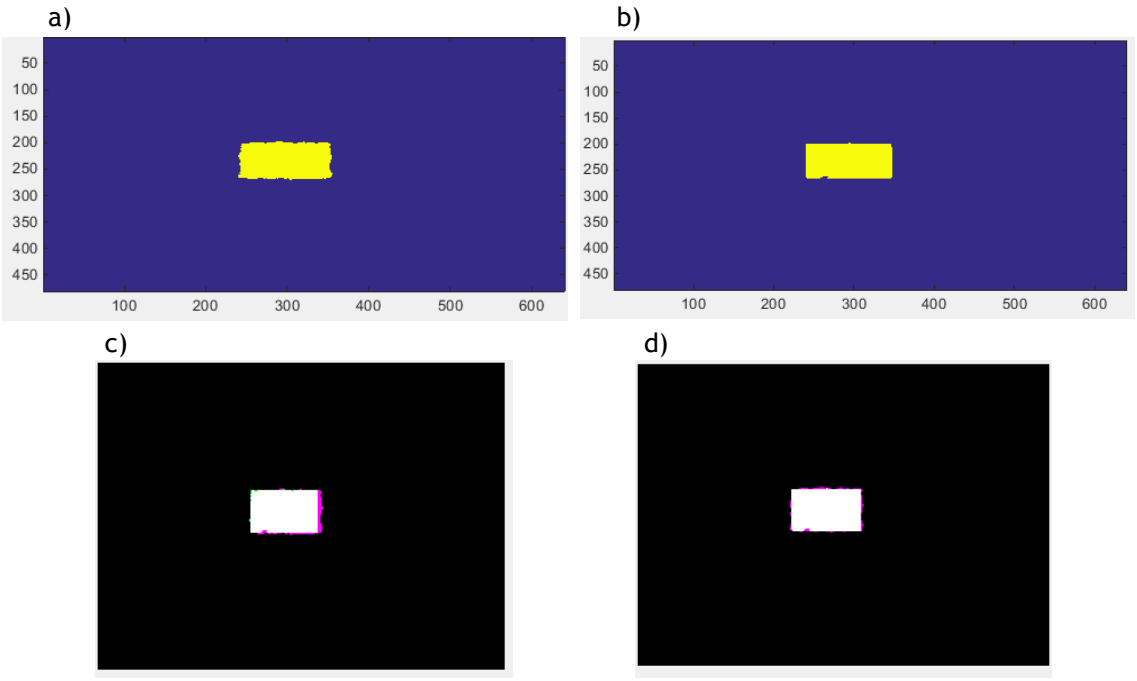


Figure 38. Results obtained after segmentation of depth and RGB images as well as before and after registration. a) and b) are, respectively, the depth and RGB masks, after segmentation. c) and d) are the combination of RGB and depth images before and after registration, respectively. The images are not at the same scale.

The obtained geometric transformation that registers both RGB and depth images, as shown in Figure 38 d), is as follows:

$$v_{RGB-depth} = \begin{pmatrix} -3.78 \\ -1.51 \end{pmatrix}$$

The order of magnitude of this value is consistent with Figure 38 c), where it is possible to observe a spatial lag between RGB and depth images and the latter is positively displaced under *x* and *y* axis. This alignment vector was subsequently used to correct the depth values prior to the implementation of the remaining methods.

3.3.3 Depth-data recovering

With the purpose of recovering the unknown depth values, a filling hole method was proposed and described in the Methodology section. The results before and after the implementation of this method, for two different sets of frames, are shown in Figure 39.

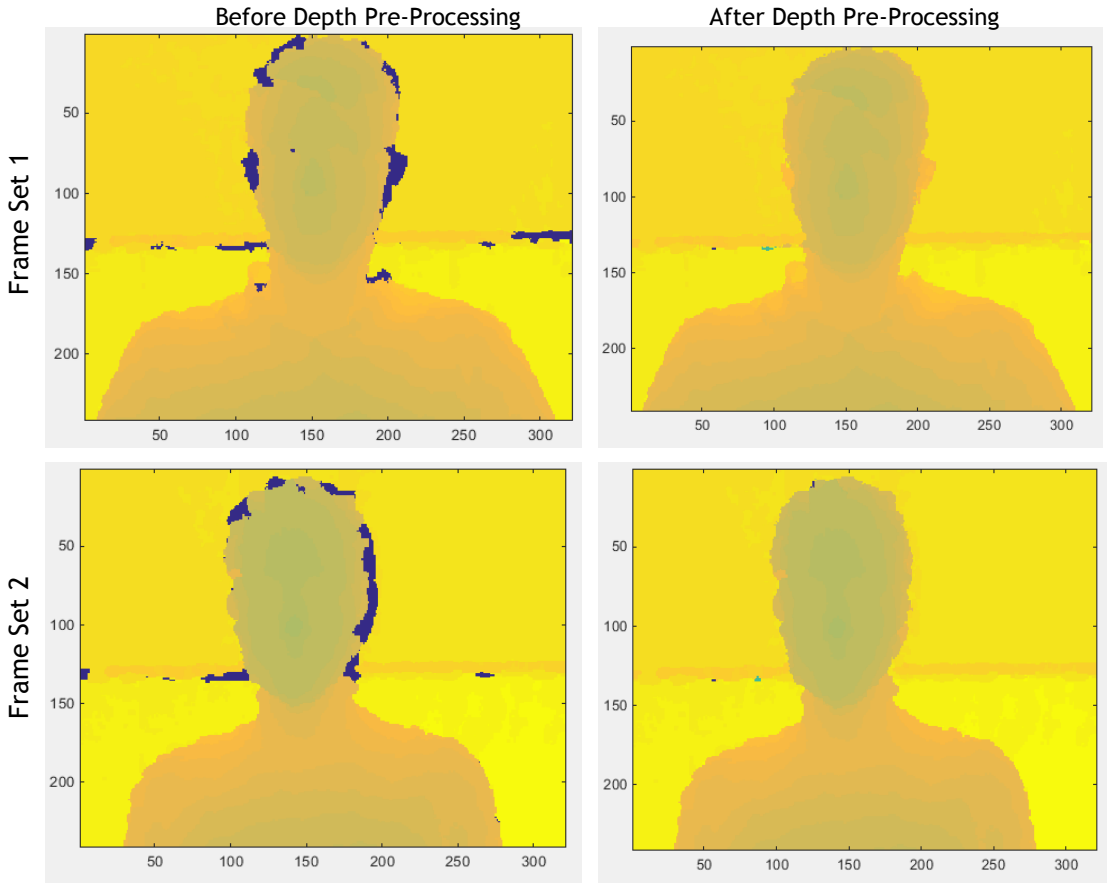


Figure 39. Depth images before and after pre-processing, for two different sets of frames. The dark blue regions represent the areas of unknown depth.

The method proved to be successful in recovering the unknown depth data, being able to estimate the missing values taking into consideration the information from previous and subsequent frames. Although the estimated depth values are, in general, closer to the theoretically ones with missing depth, when comparing to literature methods, the main drawback of the present technique is the fact that, in certain sets of frames, some pixels may not see their depth value recovered, if are pixels with unknown depth along all captured frames. Nevertheless, since these pixels correspond to object boundary pixels, such as the head or body, additional errors are not introduced during tracking due to the fact that they are not face points.

3.3.4 Head Movement Correction

The correction of the head movement is necessary to ensure that the trajectories of anatomical points and, mainly, the quantitative analysis on *facegram*, are not affected by errors. This correction was tested in a control image set (control 2), which was selected by precisely be an example of exaggerated and very broad movements. Thus, after determining the rotation and translation matrices specific for each frame, the transformation was performed to obtain the new coordinates of the tracked points.

The success of the method depends essentially on the new position of reference markers. In the case that the patient's head remains stationary throughout the acquisition, the position of the reference points is the same, with slight fluctuations, over the frames. If the face movement is accompanied by a wide head movement, after correcting it the new positions of

facial landmarks should also remain the same over time. The results in Figure 40 confirm this. In this plot, the *x* and *y* coordinates of two reference points are stable, with some local fluctuations. Therefore, these results prove the success of the implemented method.

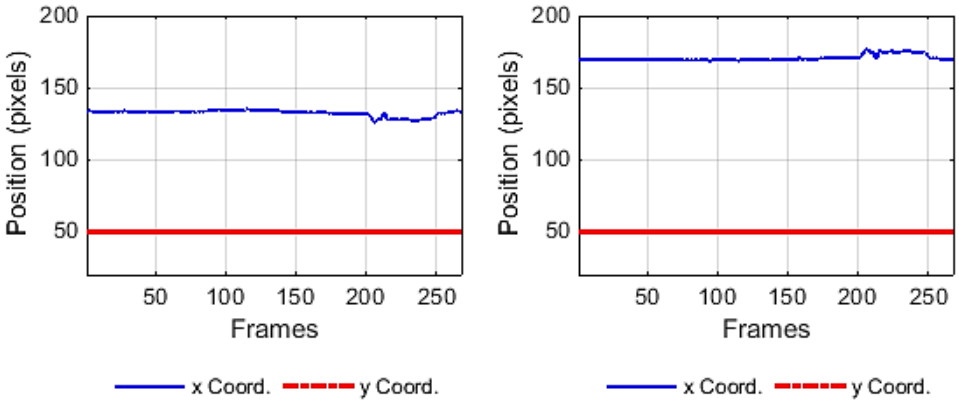


Figure 40. Graphical representations of reference points trajectories for control 2: evolution of *x* and *y* coordinates (pixels) over frames.

Additionally, an image sequence was constructed to dynamically prove the efficiency of this method. Each image underwent translation and rotation specific for each frame and the new points obtained after implement the movement correction algorithm were plotted together. These transformations are visible by the orientation of the reference line located behind the subject. The results are shown in Figure 41. The first two images relate to the original image and the transformed image, respectively, in the frame 1 with corresponding locations of facial landmarks before and after corrected. In this case, as frame 1 corresponds to a starting position without movement, both images are similar, and with no apparent correction (both reference lines are parallel). The second and third pairs of images concern, once again, the original and transformed images, respectively, in frame 110 and 200, as well as landmarks' locations before and after correction. These frames were chosen to be representative of moving situations and, in this case, it is possible to observe that the new reference points remain aligned with the ones in frame 1 and in identical positions (reference lines non collinear, due to image transformations). The transformed images show a vertical position of the subject's head, similar to the initial frame, confirming the correct adjustment of head movement.

Figure 41. Original and transformed images of frames 1, 110 and 200. Original and corrected tracked reference points represented as circles: in blue, marker 1 (Ref.1), in red, marker 2 (Ref.2).

3.3.5 Validation of the algorithms

The selection of the best techniques to perform tracking of anatomical points was achieved analyzing the performance of the tracking algorithm. To do so, the comparison between theoretical trajectories, obtained from the two control subjects, and the tracked points, resulting from the implementation of each technique, was carried out, using R^2 and mean error measurements for the purpose.

In subjects without pathologies, the anatomical landmark locations with greatest extension of motion are the commissures' points (right or left). Therefore, in order to visually observe the success of tracking and its errors regarding higher amplitude movements, control 2 was chosen to exemplify algorithm's performance and, by way of example, the obtained trajectories for the reference markers (markers 1 and 2) and marker 7 (left commissure) are illustrated in Figure 42. However, the conclusions were drawn analyzing all markers, for both controls. This example concerns the implementation of the best set of methods that performs facial points' tracking. The validation results for each method are described in the following sections.

For the purposes of scientific objectivity, it will be presented then an isolated example in which the tracking did not go as expected, for one set of images of control 1; however, in the vast majority of situations, this was totally avoided, with very promising results.

From Figure 42, an illustration of what happens in the remaining markers for control 2, it is possible to realize that, in general, the tracking was successfully completed, since it properly follows the movement of the *x* and *y* coordinates along the theoretical trajectory. The position, in pixels, of each marker, varied significantly, with maximum extents greater than 60 pixels, which underlines the fact that the facial movement in control 2 was quite wide and exaggerated.

However, the situation is different for points' tracking using images from control 1. Figure 43 illustrates a single example of algorithm failure that led to a poor performance, in the set of videos obtained for control 1. For marker 3, at some point of the facial motion, the *y* coordinate was mistakenly tracked what triggered a shift in *x* coordinate's position, leading to tracking errors, which remained until the end of analysis. For marker 5, only the *x* coordinate suffered negative deviation from the ideal position. These two examples can explain what frequently occurred in tracking facial points of control 1. When affording the smile, some areas with intensity and depth informations similar to those of markers appear in the image, such as the dynamic wrinkles on smile line above the mouth, or inside the mouth at its corners, lead to confusion when determining the point that minimizes the error in the BMA algorithm. In addition to that, the fact that the RGB information varies over frames for the same marker, due to its position and lighting changes, and as the template is always related to the initial frame, leads to the appearance of regions with more similar information to the template, when compared to the real areas, minimizing the error and being mistakenly detected as a new position for the point to be tracked. Despite these limitations, both controls were used to test the performance of the several methods, shown below.

It should be noted that markers 1 and 2 (Ref.1 and 2), in general, exhibited lower performance results, visually described by theoretical trajectories, as the examples presented above. This relates to the fact that as the position changes of the markers are very fine, any discrepancy of 1 or 2 pixels (which is what is often observed) causes abrupt alterations that may seem exaggerated regarding graphics' scales.

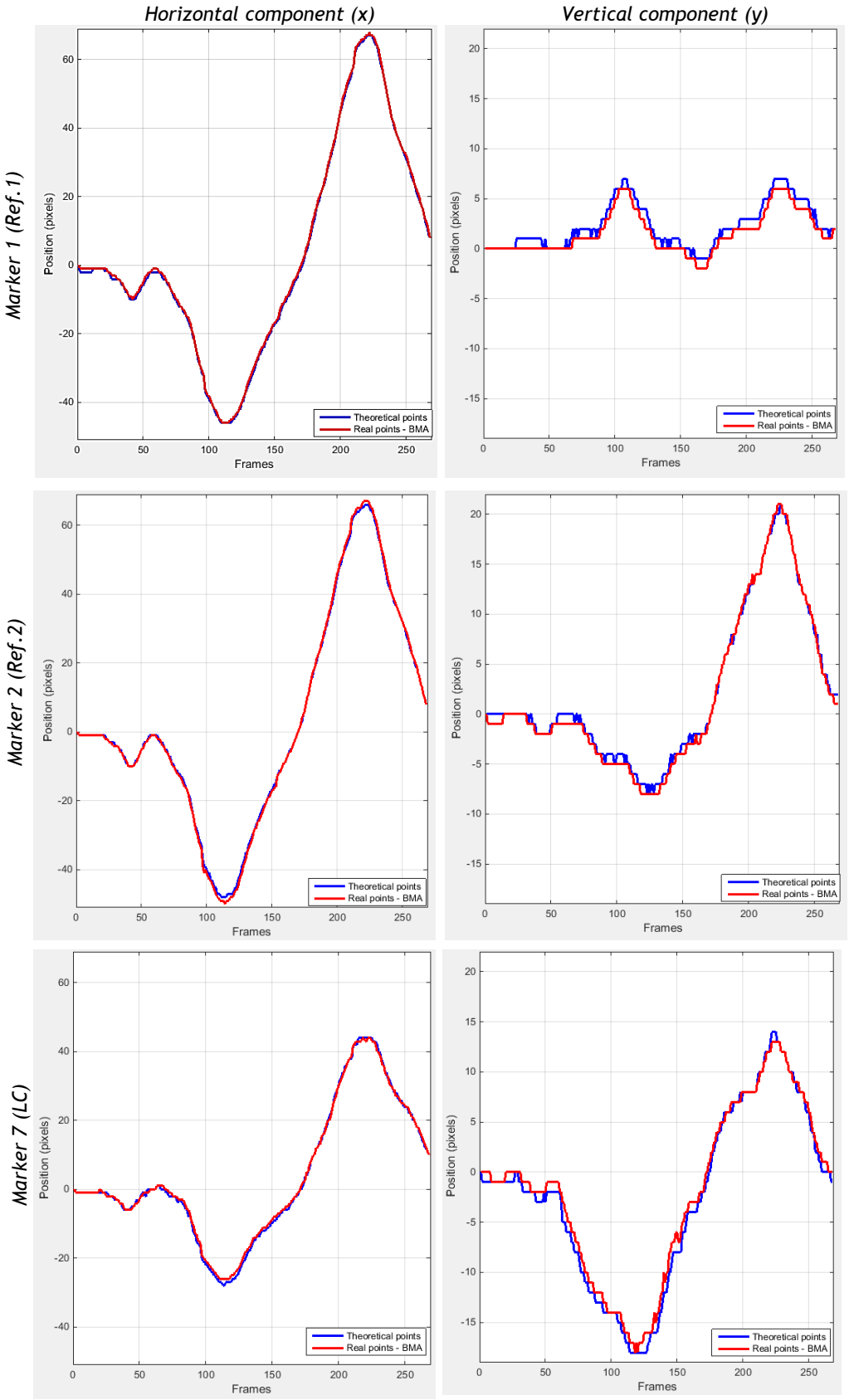


Figure 42. Theoretical (blue) and real (red) trajectories for markers 1, 2 and 7, using L1 = 7 and L2 = 11, for both horizontal and vertical components (control 2). Ref.1 = Reference marker 1; Ref.2 = Reference marker 2; LC = left commissure.

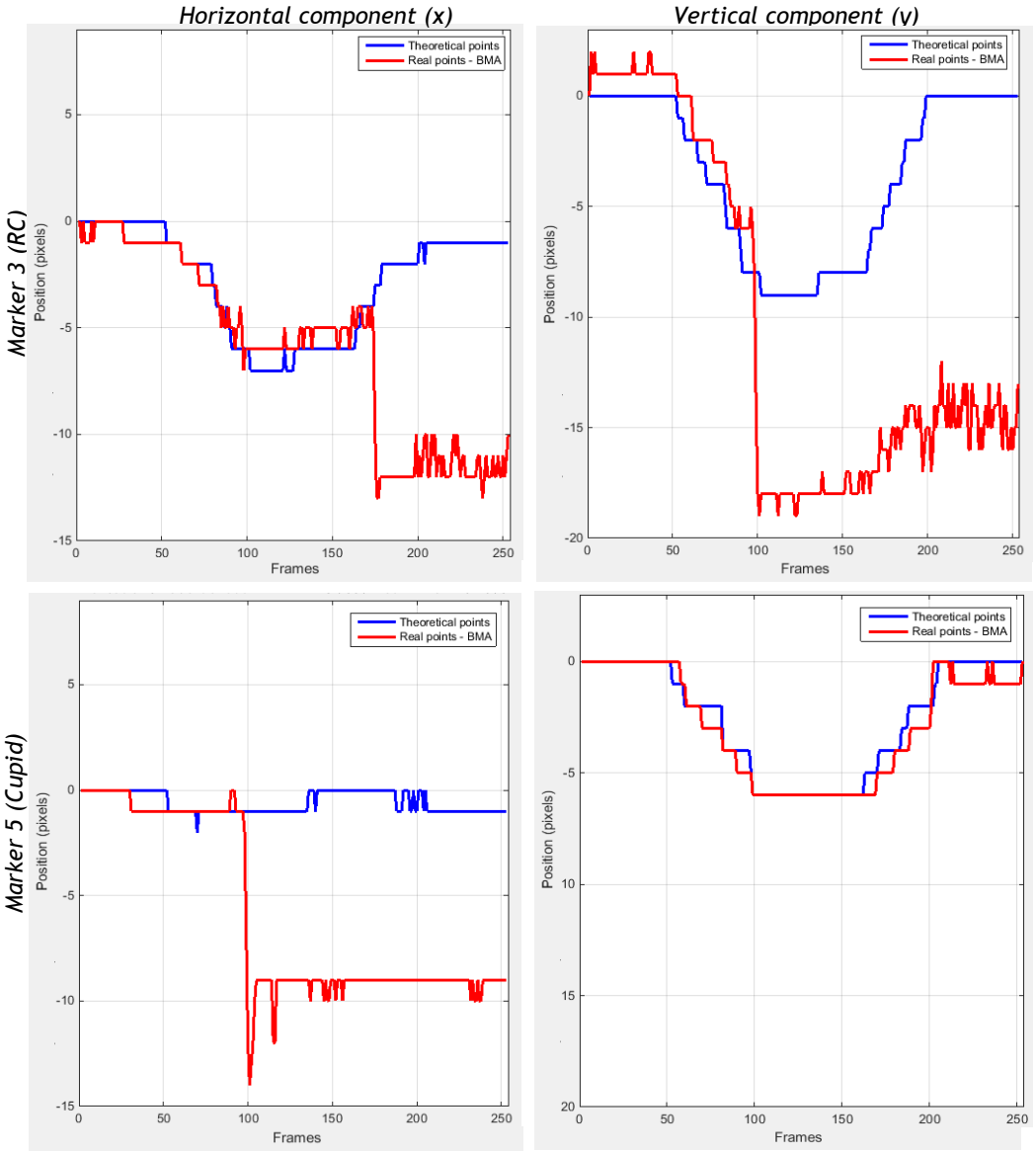


Figure 43. Theoretical (blue) and real (red) trajectories for markers 3 and 5, using L1 = 7 and L2 = 11, for both horizontal and vertical components (control 1). RC = right commissure.

3.3.5.1 Color-map contrast enhancement

Four different methods of pre-processing were tested with regard to their ability to maximize the local contrast and efficiently perform tracking: using median (with T=0 or T=1) to calculate the reference values of R, G and B; complement the above with the *LoG*; using mean for the same purpose, with T=1 and, the last one, filtering only the red channel (without additional processing). The obtained results of R^2 and mean error, for each technique and control are detailed in Table C1, in Appendix C.

Regarding control 2, which gets better results for the reasons mentioned above, it is possible to observe that, in general, the technique that uses the average together with *LoG*, for T=1, revealed better results, i.e., higher values of R^2 and lower mean error, with significant improvements in the case of the marker 1. In the cases where it does not happen, the differences regarding what was considered the best are in the order of hundredths, for both quality measurements.

In case of control 2, the results are in accordance with the previous; however there is greater discrepancy between the best results. While in control 1, the best results were often obtained using median technique with T=0, in this control, filtering the red channel obtained the second best results. What happens when only the red channel is filtered is that it works well for, for example, black color markers, which was the color used in color markers for both controls and patients, limiting the marker color to be used. Therefore, this technique was not selected to include the best set of methods to enable versatility in choosing the marker's color.

The fact that, in certain markers, the results obtained for the technique that uses T=0 were the best can be explained as follows: as intensity values around the center point are very different from each other, the mean or median will introduce errors that will extend over the technique. When using only the center pixel of each marker, which in theory is well chosen by user as being the pixel that has the real color of the marker, the comparison between the pixels of the remaining image is made relative to this one, and the contrast is increased (this conclusions are supported by contrast results, presented below). Since it is not correct to restrict the analysis to only the user-selected pixel, in order to avoid tracking errors if it is incorrectly chosen, the technique that uses the median with T=0 was discarded. Accordingly, the use of median along with *LoG* filter and T= 1 was the chosen approach to perform pre-processing due to its successful results.

To supplement the analysis of tracking's performance and the choice of the best set of methods, the contrast was calculated using MSE for 3 methods: filtering only the red channel, using the average or the median for calculating values of R, G and B reference. The contrast results are shown in Table IV. These are supported by the explanation given above, which indicated why the method that uses the median and T=0 is the one that allows higher contrast results.

Table IV. Contrast results for images from both controls (first frame only)

Method	*Control 1*	*Control 2*
Red Channel	728.76	2.52×10^3
Median T=0	1.94×10^3	3.65×10^3
Median T=1	770.97	1.40×10^3
Mean T=1	571.91	1.20×10^3

3.3.5.2 Tracking of facial points

To understand how the size of template (L1) and the search block (L2) affect point's tracking, the performance of algorithms was tested for four different sets of sizes:

Set 1: L1 = 11 and L2 = 13;
Set 2: L1 = 13 and L2 = 21;
Set 3: L1 = 11 and L2 = 21, and finally,
Set 4: L1 = 7 and L2 = 11.

Thus theoretical trajectories of two sets of frames were constructed (control group) and, once more, compared with the trajectories obtained after tracking (for the 7 markers individually and for each horizontal (x) and vertical (y) component).

The results regarding R^2 and mean error values for all markers and sets, as well as for both control subjects, are detailed in Table C2, in Appendix C.

For control 2, the best results regarding R^2 and mean error were generally obtained using the set 4 (*L1* and *L2* equal to 7 and 11, respectively) with R^2 values close to 1 and mean error not exceeding 0.2 pixels (approximately 0.4 mm). The derogations presented differences of the order of hundredths compared with the results obtained for the set 4.

Once again, for control 1, despite the most optimal results are concentrated on set 4, there is value dispersion over the remaining frames and the tracking was only successfully achieved for markers 6 and 7. These results can be explained by the same reasons in above section.

Thus, the results suggest that templates sizes included in set 4 are more appropriate to optimize the tracking process. In fact, it was found that the lower the sizes of the templates are, the better the behavior of algorithms; however, as mentioned above, this size reduction has a minimum value since it is necessary to take into account not only the selected point information but also the point's neighborhood to provide the right framework. Otherwise, tracking becomes limited, and the results do not vary along the facial movement.

In addition to test the influence of *L1* and *L2* sizes, two alternatives for template block construction were also tested – center the template on the tracked pixel of the previous frame or 5 frames ago. When testing the first approach (using a template block centered on the tracked point of the previous frame) it was found that, for a given marker, the coordinates over time were kept constant. This can be explained as follows: the movements to be measured are very smooth, compared to time acquisition rate (below 30 Hz) and the spatial resolution is low. As such, between consecutive frames, the location changes of the markers are of subpixel, i.e., since the difference between frames is always smaller than a pixel, it is not possible to measure such displacements.

Thus, the comparison between real and theoretical trajectories was performed only for the second approach, regarding R^2 and mean error values. In this case, the results for both control subjects are shown in Table C3, in Appendix C. These results should be compared with the ones from Table C2, in Appendix B, for set 4. Therefore, when using the template centered in the tracked point 5 frames ago, the R^2 values are more distant from 1, than using the previous frame to construct the template. In addition, the mean error values are higher, which determines a greater distancing between real and theoretical points. These results are in agreement for both controls, although for control 1 there is more dispersion of good results, as usual. In addition, the trajectories of tracked points showed local fluctuations over all frames, as compared with the theoretical trajectories (data not shown).

For this reason, the template block always centered on the first frame point was the technique used to obtain all subsequent results.

3.3.5.3 Block Matching algorithm

The minimization of the cost function used in the *BMA* for facial points' tracking was performed in 4 different ways, as explained in the methodology section. In order to test the performance of each technique, the trajectories of the points obtained using each of these

techniques were, once again, compared to theoretical trajectories obtained for control groups. Considering the results on Table C4, in Appendix C, for both controls, it is possible to verify that the greatest R^2 were obtained, in most cases, when using only the minimum as technique to minimize the cost function (first approach), followed by the method that uses the median of sum-matrix's values. Regarding now the mean error, the conclusions were the same. As a result, the developed method using the first approach appears to be a more suitable solution for facial points' tracking.

3.3.5.4 Validation in a controlled environment

After acquiring RGB and depth images of the structure used for calibration in a controlled environment (see examples in Figure 44), the tracking of the central point was performed and theoretical trajectories and real points, obtained after tracking points, were plotted together (Figure 45).

By graphical analysis, it is possible to conclude that tracked points approach the theoretical trajectory, in each segment. To quantify the performance of tracking algorithm regarding the horizontal, vertical and depth components, coefficients of determination and mean error were calculated for each section of the trajectory (numbered 1 to 8). The results, presented in Table V, help to confirm the efficiency of tracking with respect to each component, since R^2 are so close to 1 and the mean error do not exceed 3mm.

To the best of author's knowledge, there is no work in the literature similar to the one presented in this research, i.e., in which both depth and intensity information are used to perform tracking by BMA, so it is not possible to compare the performance of the developed method. In this way, it can be concluded that the method successfully performed tracking, not only for horizontal and vertical components, but also for depth, and the success of this experiment appeared to be a further proof of the performance of the tracking algorithm.

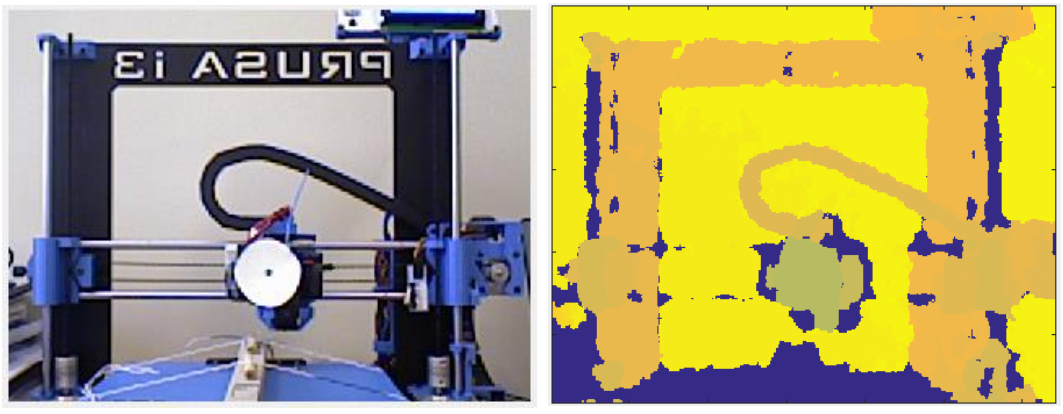


Figure 44. RGB and depth images from computer-controlled system for calibration of x, y and x coordinates (cropped images). Blue spots in the depth image represent unknown depth data. Distance from the camera: 0.8m.

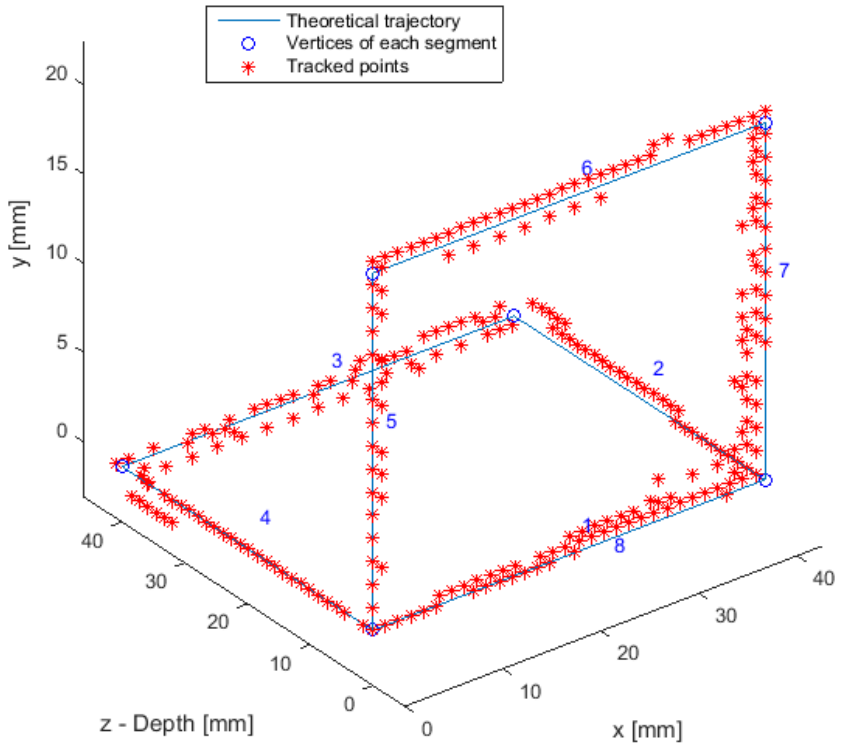


Figure 45. Theoretical trajectory (blue) with vertices of each segment (blue circles) and real points obtained after tracking (red asterisks) as function of *x*, *y* and *z* coordinates. Distances converted to millimeters.

Table V. Determination coefficients and mean error obtained after comparing the theoretical trajectory with tracked points for each section of the trajectory.

Section	1	2	3	4	5	6	7	8
R^2	0.973	0.985	0.982	0.998	0.989	0.996	0.925	0.996
Mean Error [mm]	2.87	2.00	2.63	0.246	0.354	0.57	2.48	0.660

3.3.6 Trajectories of facial points

3.3.6.1 Construction of facial points trajectories

The tracked points obtained using the tracking algorithm, without any data processing, allowed the construction of two types of graphic trajectories, previously described.

Examples of type 1 trajectories are illustrated in Figure 46 for control 1. From these initial tracings conclusions can be visually drawn regarding the evolution of markers positions for each vertical and horizontal component separately, as well as coordination between *x* and *y* coordinates and markers. It should be noted that, since these data were not processed or filtered, the information that can be drawn from these graphics is very basic and preliminary.

In the example shown, for a subject without pathological conditions, markers 1 and 2 show a constant trajectory over time, with 1/2 pixel's fluctuations. Markers 3 and 5, as mentioned above, were wrongly tracked, as can be seen by their trajectories. In the case of

the marker 4, both components have an apparent coordination over time with progressive distancing and approaching along the facial movements. This is the behavior that was expected to observe for marker 3. One can still realize, through the *x* coordinate trajectory, that the initial and final position of the marker was not the same. Finally, markers 6 and 7 showed similar behavior, with apparent temporal coordination and stabilization of the coordinates in rest and hold intervals.

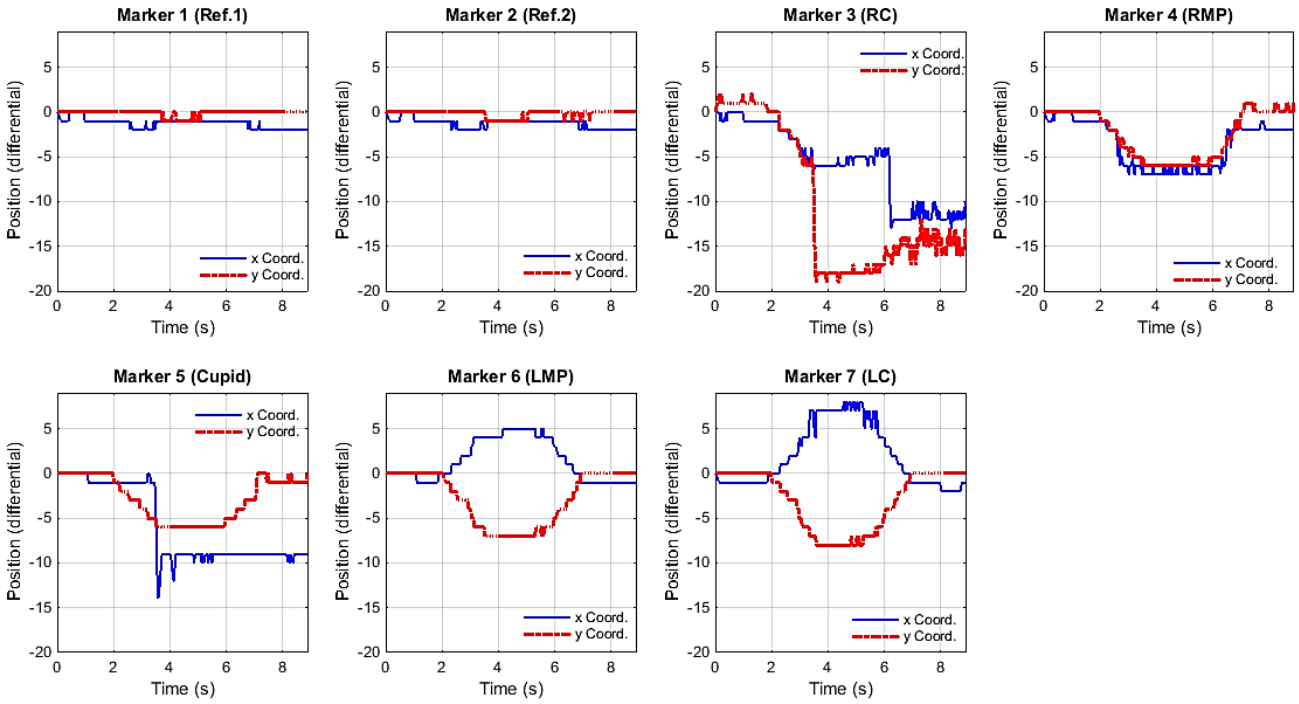


Figure 46. Graphical representations of anatomical points' landmarks for a normal subject (control 1): evolution of *x* and *y* coordinates (pixels) over time (type 1). Ref.1 = Reference marker 1; Ref.2 = Reference marker 2; RC = right commissure; RMP = right midpoint; LMP = left midpoint; LC = left commissure.

The depth information may also be graphed in order to conclude about depth evolution over facial movements for each marker. In Figure 47, examples of depth trajectories are illustrated examples, for control 1.

In a normal subject without pathologies and regarding the 5 anatomical points, it is expected that during contraction the depth increases its value, which remains constant during the hold phase, decreasing again until the initial resting position. For the reference markers (Ref.1 and 2), the depth value must remain approximately constant across frames.

In fact, and as illustrated by Figure 47, the depth values obtained for the reference markers continuously oscillated around the starting point, with fluctuations characteristic of the depth measurements. For markers 4, 6 and 7 the obtained depth trajectories were as expected. However, there was no such agreement for markers 3 and 5, whose explanation has been given in the previous sections.

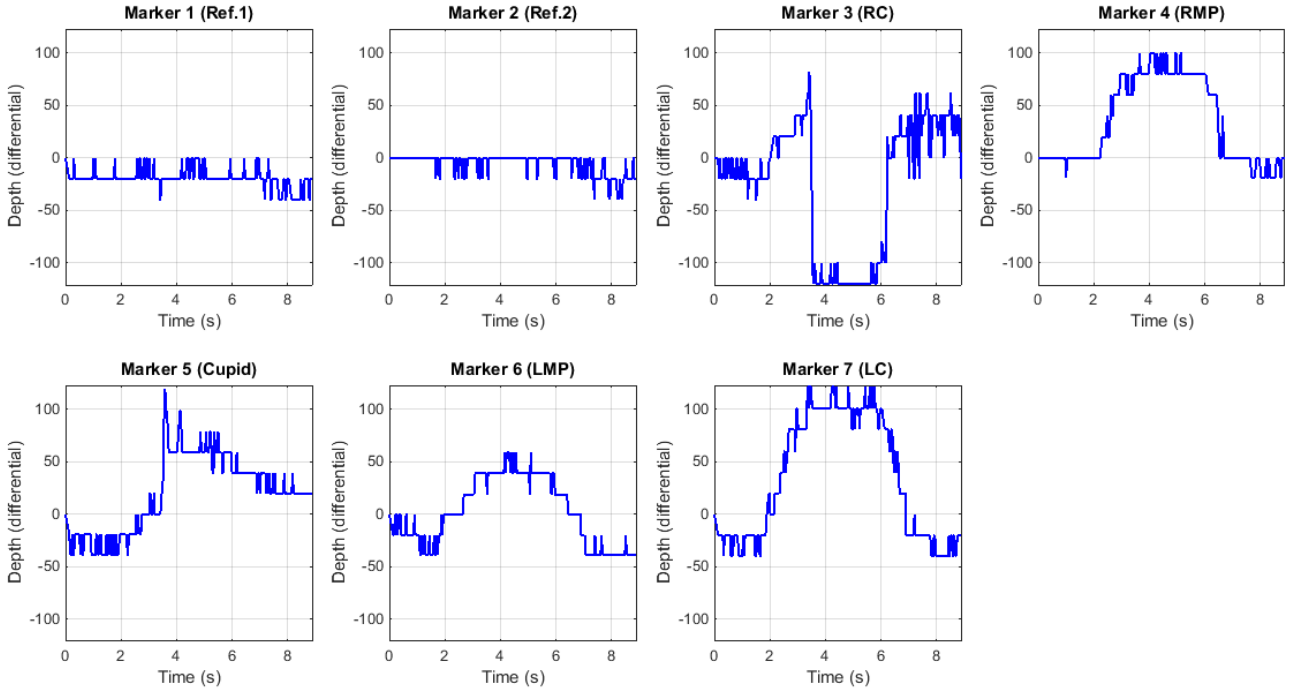


Figure 47. Graphical representations of anatomical points' landmarks for a normal subject (control 1): evolution of depth values over time (type 1). Ref.1 = Reference marker 1; Ref.2 = Reference marker 2; RC = right commissure; RMP = right midpoint; LMP = left midpoint; LC = left commissure.

Trajectories of anatomical points from patients 1 and 2 were also analyzed concerning the evolution of *x* and *y* coordinates and depth values over time. In Figure 48 are presented the markers' trajectories of patient 1.

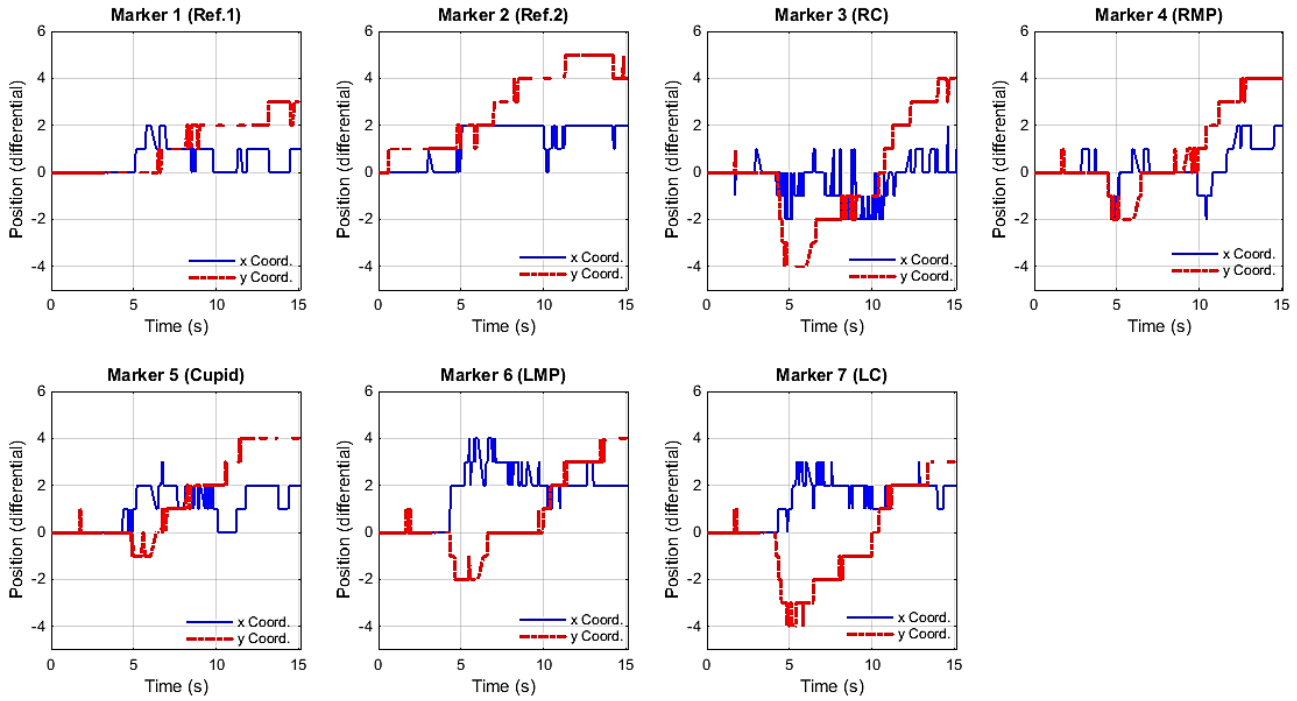


Figure 48. Graphical representations of anatomical points' landmarks for patient 1: evolution of *x* and *y* coordinates (pixels) over time (type 1). Ref.1 = Reference marker 1; Ref.2 = Reference marker 2; RC = right commissure; RMP = right midpoint; LMP = left midpoint; LC = left commissure.

In this patient, it is possible to conclude that initial and final positions of all markers are not the similar, suggesting the presence of contraction or partial relaxation in the final phase of acquisition, rest phase. Trajectories of markers 3 and 4 propose that right commissure and midpoint movements are almost exclusively horizontal. The left side shows greater amplitude of excursion in both components (horizontal and vertical). However, the data must be smoothed and filtered to eliminate the local variations but keep possible sudden increases which may indicate the presence of spasms and synkinesis.

The same happens in depth trajectories (Figure 49), with lower extensions on the right side of the face. Once again, it becomes necessary to filter the data in order to enable a more accurate and assertive analysis.

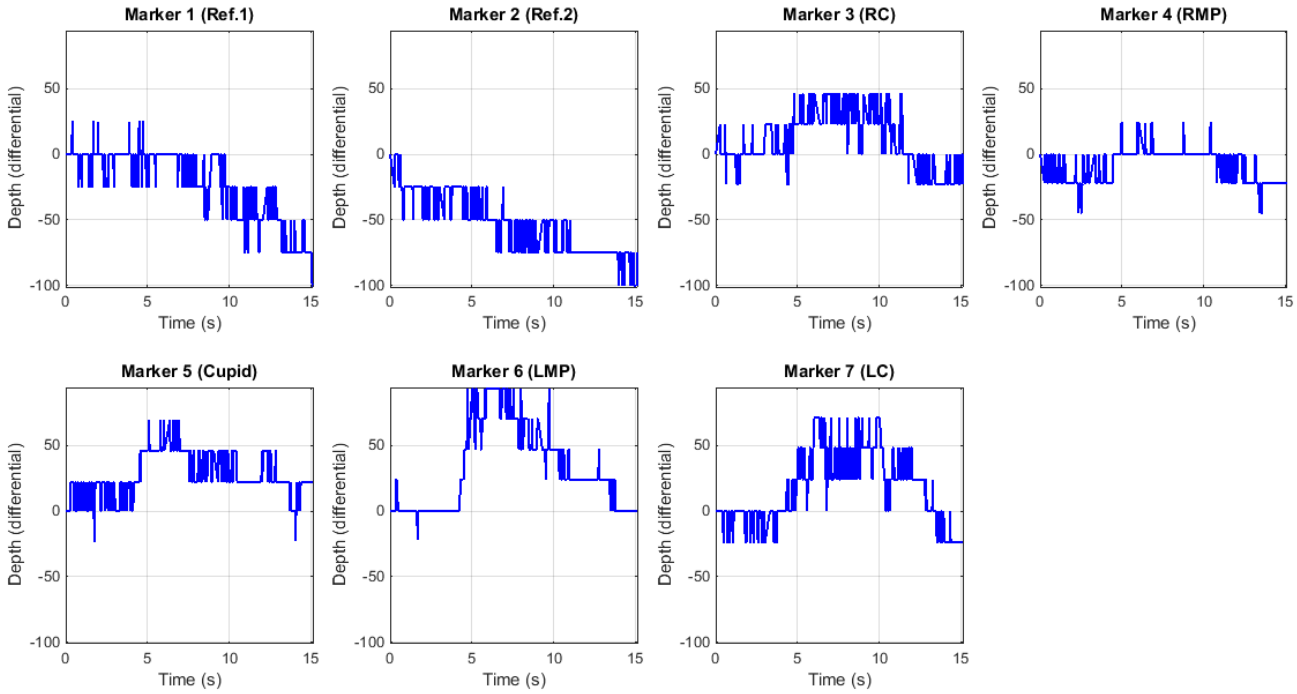


Figure 49. Graphical representations of anatomical points' landmarks for patient 1: evolution of depth values over time (type 1). Ref.1 = Reference marker 1; Ref.2 = Reference marker 2; RC = right commissure; RMP = right midpoint; LMP = left midpoint; LC = left commissure.

Examples of the second type of graphic, or parametric trajectories, are illustrated in Figure 50, for control 1. By the way of example, markers 4 and 6, respectively, left and right midpoints were chosen to exemplify these tracings.

In addition to describe the markers' trajectories so that, visually, it is possible to associate the patient's motion to markers' excursions represented on plots, these graphics allow identifying asymmetries in relaxation and contraction paths.

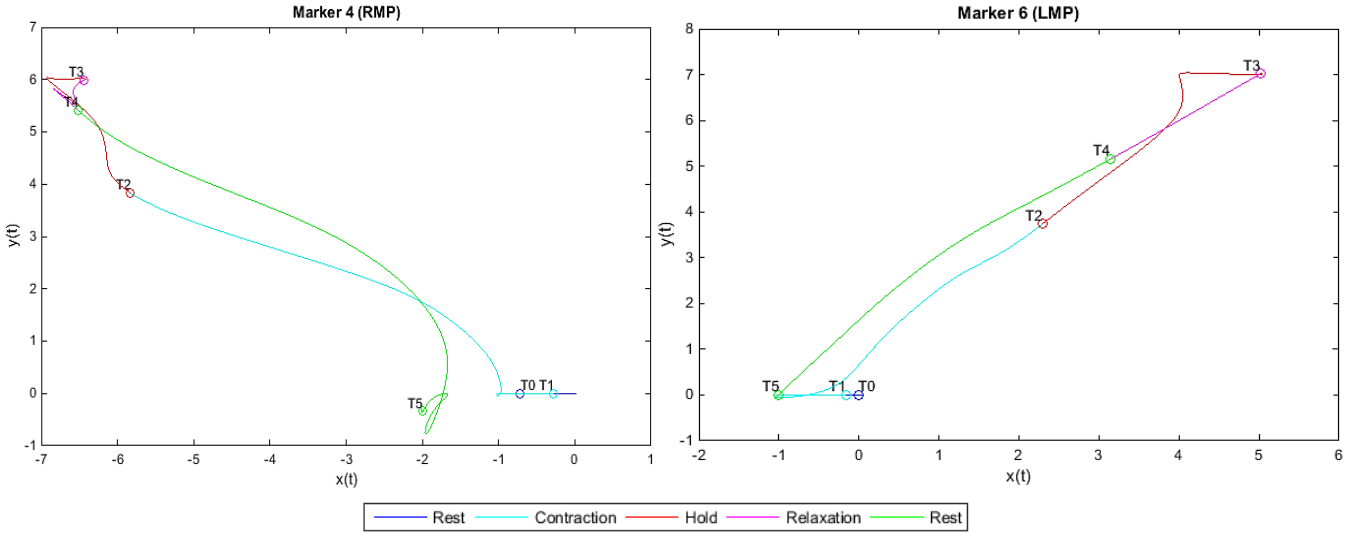


Figure 50. Graphical representations of pathways for markers 4 and 6, where 6 points were mark important times: start of recording (T0), start of contraction (T1), end of contraction (T2), start of relaxation (T3), end of relaxation (T4) and end of recording (T5). Control 1.

A typical excursion of a subject without pathologies must be represented by a smooth line through time. In fact, the paths shown above describe an approximately linear behavior between rest phases, with small variations.

In the case of patients with pathologies, such as Patient 1 (Figure 51), the excursion is recorded in a non-smooth line with several inversions and abrupt direction changes. Either way, it is possible to observe that marker 6 (left midpoint) shows greater amplitude of motion, and the trajectory is more similar to an excursion normal. For both, it can be seen that the initial and final points do not coincide, as described above.

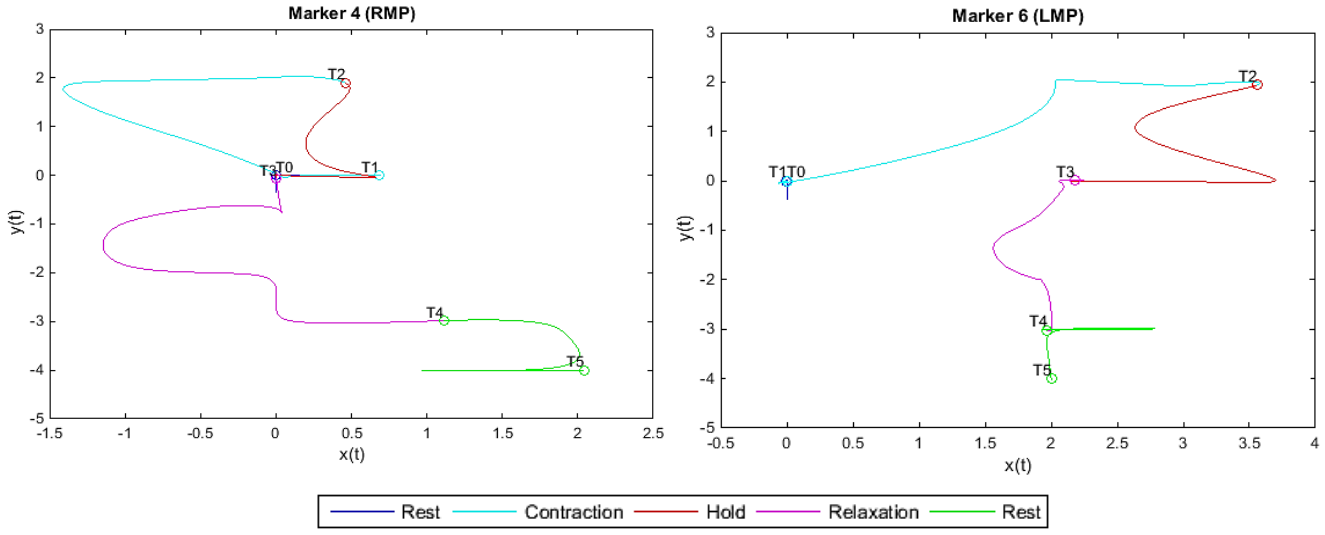


Figure 51. Graphical representations of pathways for markers 4 and 6, where 6 points were mark important times: start of recording (T0), start of contraction (T1), end of contraction (T2), start of relaxation (T3), end of relaxation (T4) and end of recording (T5). Patient 1.

For the construction of these trajectories, it is necessary to interpolate data to obtain a curve, continuous on time. Examples of obtained interpolated curves are illustrated in Figure 52. As intended, the tracked data were filtered by removing disparate points, resulting in a smooth curve that better approximates the trajectory of each marker.

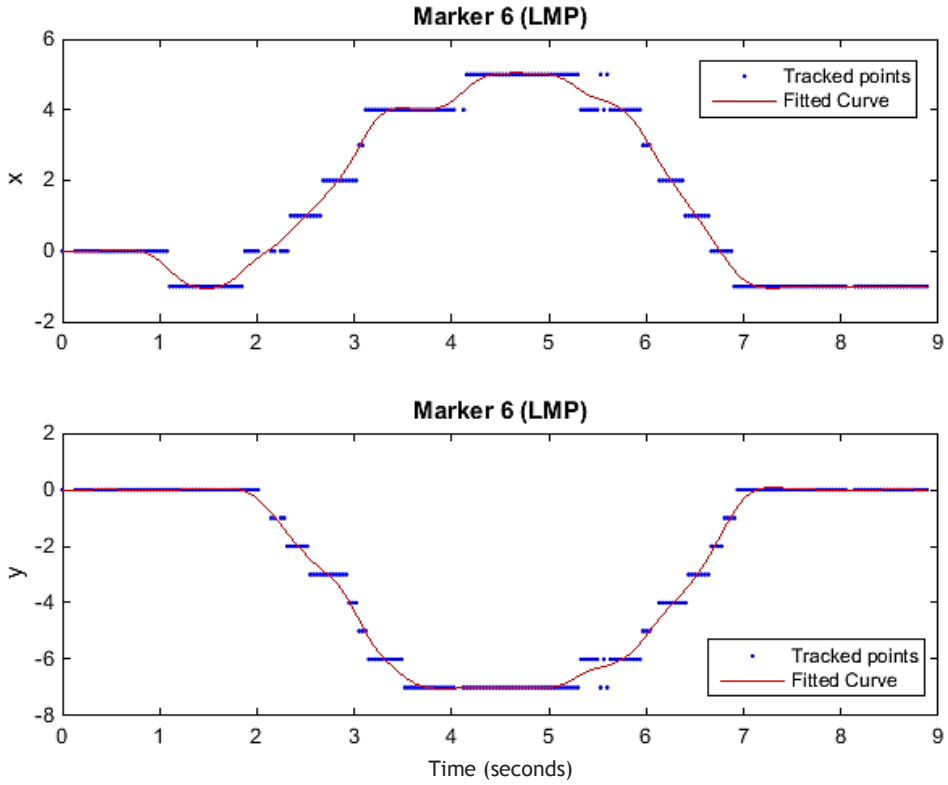


Figure 52. *x* and *y* coordinates (above and below, respectively) as function of time (in seconds) (blue dots), interpolated with a smoothing spline (red line) (*p*=0.99).

3.4 Conclusion

In order to create a potential system for the quantification of facial anatomical points' motion through tracking of their positions over time, a set of methods using a low-cost equipment (a *Kinect*-based camera) was presented in this chapter. In fact, this type of camera is required since facial movements are of extreme complexity and can only be fully characterized using 3D analysis.

The first stage consisted in collecting RGB and depth images using the *Asus Xtion Pro Live* camera (a RGB-D sensor), for further processing and analysis. Besides the collection of 3D data from two subjects with no apparent pathology, RGB and depth images were captured on clinical environment, from two patients, using an application for medical data acquisition.

To ensure the necessary initial conditions for a good tracking and quantitative analysis of facial points' motion, spatial and temporal depth precision was tested and a method that allows depth-color registration was implemented. Consequently, RGB and depth preprocessing was achieved using an algorithm for removing the unknown depth data and obtain images with improved contrast and edges, respectively. Then, an effective tracking of facial points was succeeded, using *Block Matching Algorithm* to match points between

different frames. This method combines RGB and depth information, to achieve better results comparing to other available methods. In order to safeguard potential fluctuations in patient's head during data acquisition, a method for correction of these movements was implemented and tested, ensuring accurate determination of point's positions.

Using the trajectories of points whose positions were manually determined, the possibilities for the different implemented methods were explored and tested regarding tracking's performance, aiming to determine the set of methods that maximize the tracking. The best results, i.e., the methods that achieved greater values of R^2 and lower mean error values, were obtained using the following group of methods: median with T=1 complemented with *LoG* filter (pre-processing algorithm); L1 and L2 equal to 7 and 11, respectively; the minimum of the sum of the errors in RGB and depth to calculate the respective point with *BMA* and template block should be centered on points from first frame (tracking algorithm). In addition, the precision of tracking regarding depth component was analyzed using a controlled environment thereby obtaining determination coefficients close to 1, confirming the successful tracking. Finally, the trajectories of all markers were constructed using two types of tracings and the results for control 1 and patient 1 were discussed. As a matter of scientific disclosure, a particular case in which the tracking algorithms did not have the desired behavior was presented. However, this case was an exception, resulting from the onset of areas with most similar color to the template block, but far away from the marker's position, in the tracking algorithm.

As previously mentioned, the available methods for quantification of facial movements do not allow a complete characterization, either lacking of 3D information, having high computational time or very expensive equipment and complex setups. Currently, qualitative methods are typically used, but, because they are subjective and depend on whether patient's or physician's opinion, are unsuitable for an objective and quantitative characterization. Thus, it is of utmost importance to create a quantitative, universal and inexpensive method for assessing treatment's outcomes and assist doctors, providing clinically useful information.

The proposed method can be a possible solution for 3D quantitative characterization of facial movements, using low-cost equipment with high performance at a low computational cost.

Chapter 4

Facegram

The results of interventions in reconstructive plastic surgery of the face are currently analyzed using qualitative techniques and, as explained in the previous sections, these methods are insufficient to quantify the mobility degree of patient's face. In fact, the visual observation by physicians is not always enough to assess these surgical outcomes, even less the patient or clinician's opinion, and there are no standardized methods for this evaluation.

With this in mind, the last goal of this research is to create an interface that includes a set of distilled measurements with clinical usefulness and that can be easily interpreted by a physician. This set of morphological measures, whether dynamic or static, proposed to be named *facegram*, is complemented with results obtained from the projected system described above. With this tool, the author believes that it will be possible to quantitatively assess the normal or pathological behavior of the set of anatomical landmarks, while being simple, inexpensive and quite functional. In fact, the selection of the best measurements that allow this quantitatively assessment was performed with the help of a plastic surgeon.

In this way, the entire content of *facegram* was designed to include only relevant clinical information and easily accessible for quantification of facial movement. The organization of the *facegram* will be described in the next section.

4.1 *Facegram* Components

The contents of *facegram* will be displayed in the form of a report, with PDF format, taking advantage of *MATLAB Report Generator Toolbox*[3] to generate the template. Initial information such as name, age, sex and date will be firstly displayed, as well as a representative frame of each movement's phase (and respective duration in seconds).

The analysis and quantification of movement is presented later and is divided into two different categories: static and dynamic analysis.

[3] http://www.mathworks.com/products/ML_reportgenerator/

4.1.1 Static analysis

All graphical representations and measurements were constructed taking into account the information obtained about the position of the tracked points over time.

The first step is to transform the obtained coordinates in values of medical interest, which can be graphically represented in an easy way with clinical usefulness.

The horizontal components (*x* coordinates) of the various points (right and left commissures, right and left midpoints and cupid) were calculated using Equation 21, which represents the distance from a point to a line in 2D. In this particular case, each *x* coordinate is the distance between the corresponding tracked point ($p_{x,y}$) and the symmetry axis, characterized by the two points ($t1_{x,y}$ and $t2_{x,y}$), previously defined in the application for tracking and visualization of medical data (described in section 3.2.6).

$$d = \frac{\left| \det\left(t2_{x,y} - t1_{x,y} \;\; t1_{x,y} - p_{x,y} \right) \right|}{\left| t2_{x,y} - t1_{x,y} \right|} \tag{21}$$

where *det* denotes the determinant of $\left[t2_{x,y} - t1_{x,y} \;\; t1_{x,y} - p_{x,y} \right]$ matrix.

In turn, the y coordinate of each interest point of interest was calculated as the distance to cupid at frame 1, only relative to vertical component, i.e., the difference between the *y* coordinates of tracked point and cupid. The *z* coordinate, depth, is calculated in the same way and, in this case, the reference value is the depth of cupid, also in frame 1.

Since all computations were done in pixel-space and in order to create a clinically useful and easy-to-interpret tool, the positions need to be converted to real-space (millimeters). This can be done by calculating the scaling factor (pixel/mm) obtained from the distance (in millimeters) between the two reference markers, initially provided by the user, and the distance in pixels of the same markers in frame 1. Instead, and as explained in section XX346, this scaling factor can be obtained from camera calibration, without needing to calculate it for each specific case.

In this research, the scaling factor that was used to convert the measurements to real-space (only horizontal and vertical components) was calculated for each subject with the distance between the markers, previously measured in mm. In fact, this conversion can only be done assuming that the depth band is so short that the scale factor remains unchanged for the analyzed points. Depth measurements were converted to real-space by a scaling factor equal to 0.1, since depth raw data are given in tenths of a millimeter.

The first graphical representation that appears on *facegram* regarding static analysis is the 2D tracing of each anatomical point's position at rest (intermediate frame of *Rest* phase) and it was constructed using the *x* and *y* coordinates calculated as previously mentioned.
Similarly, other tracing outlined in static analysis section corresponds to the position's representation of the diverse markers but, this time, at one of the frames corresponding to the supposed maximum extension (first frame of *Hold* phase). Both graphical representations allow the analysis of the position of anatomical landmarks at rest and maximum extension in order to visually assess the resting symmetry and symmetry on maximum extension movements, between different points. In fact, qualitative methods are especially focused on these two facial features [13, 14, 18], having great clinical utility, and for this reason it was given due importance in this section.

To complete this graphical information, maximum extensions in millimeters for each marker were displayed on *facegram* in table form. These vales correspond to the 3D distance

between an initial point and the same point on the maximum extension frame. It should be noted that the maximum extension frame is specific to each marker, not necessarily being the same for all facial points.

Finally, in order to create a quantitative measure that is solid enough to give useful information to physicians and that summarizes the static analysis, a morphological measure was created and named symmetry index. This static measure is intended to provide information about the relative position of points for each pair of anatomical landmark, right and left commissures pair and the other pair of midpoints, and it was calculated using Equation 22. Figure 53 represents a schematic illustration of the different variables used to calculate symmetry index, in 3D.

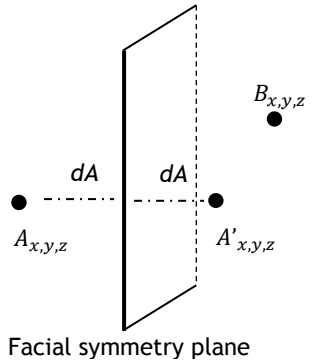


Figure 53. Representation of useful variables, used to calculate symmetry index. *A* and *B* represent the pair of analysed markers; *A'* is the projected point of marker *A* along the plane of symmetry; *dA* is the distance between A and the facial symmetric plane, calculated using Equation 22.

$$symmetry\ index = \frac{1}{1+\frac{\sqrt{(A'_x - B_x)^2 + (A'_y - B_y)^2 + (A'_z - B_z)^2}}{dA}} \quad (22)$$

where A'_x and $A'_y$are ***x*** and ***y*** coordinates of marker A', respectively and $(A'_x - B_x)$, $(A'_y - B_y)$ and $(A'_z - B_z)$ are the absolute differences between ***x***, ***y*** and ***z*** coordinates of A' and ***B***, respectively.

This quantitative measurement, within the range 0 to 1, is mathematically interpreted as the sum of errors in each direction (horizontal, vertical and depth), normalized by *dA* and recaps static analysis at maximum extension in a single number to facilitate clinical interpretation. Regarding a subject without pathologies, the symmetry index value for both marker's pairs is expect to be very close to 1, approaching zero as A' and ***B*** points are no longer coincident and with increasing distance.

4.1.2 Dynamic analysis

In addition to static features, the proposed *facegram* contains elements of dynamic analysis, essential to fully characterize and quantify facial movement.

The first 3 graphics are similar to type 1 tracings, described in section 3.2.6, regarding facial landmark's trajectories over time. The first one represents ***x*** coordinates of each marker as function of time, i.e. the anatomical landmark's distance to the symmetry axis along motion. The next one relates to ***y*** coordinates evolution, that is, the distance of each marker to cupid over time. And finally, the distance in the depth component (***z***) is also

plotted. Again, these graphics provide clinically relevant information such as the coordination between the various markers at different stages of facial movement (dynamic measure) and for each special component separately, the symmetry between each of these phases (static measure also) but visual information on velocity in every trajectory section. It is noteworthy that the curves of each marker are already interpolations of discrete points' trajectories obtained earlier, i.e., trajectories already filtered, determined as described in section 3.2.6 and with the same parameters.

The dynamic analysis is completed by tracings that illustrate the parametric paths of each pair of markers constructed with the curves described above (only ***x(t)*** and ***y(t)***), i.e., interpolated tracings of (***x(t)***, ***y(t)***) as a function of *t*. These informations allowed the construction of two plots, each referring to a pair of markers (right and left commissures, and right and left midpoints) with visual description of original trajectory of anatomical points. These graphs are similar to those described in section 3.2.6 (type 2 trajectories), this time grouped by pairs of contralateral facial points. Likewise, 6 temporal points were marked in each trajectory, corresponding to transition points between each movement phase. Information such as symmetry in the contraction and relaxation paths, movement's smoothness in each phase and visual differences between the two markers of a pair can be extracted from these plots.

It should be noted that the fact that both graphs are divided into the different stages of facial movement (5 phases), the visual analysis of the movement's dynamics is facilitated.

To more easily extract information about the velocity in each trajectory's section, a parameter that is relevant to the clinical to understand the parallel evolution between two markers of a pair, velocity tracings for each marker as a function of time were constructed. The velocity was calculated for each instant of time, taking into account both the horizontal and vertical components, as follows:

$$v = \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2} \tag{23}$$

where ***dx***, ***dy*** and ***dt*** are the differences between adjacent elements of ***x*** and ***y*** coordinates and time vector, respectively.

Finally, a quantitative and morphological dynamic measurement, called synchronization index, was determined for each pair of markers and each component (horizontal and vertical), separately. The statistical measure chosen to evaluate the correlation between two sets of data was ***Kendall's tau*** correlation coefficient, which measures the strength of dependence between two measured quantities. Considering two samples, $x_1(x, y)$ and $x_2(x, y)$, being ***x(t)*** curves for 2 contralateral markers and sample size equals to *n*, the equation that is used to calculate the value of *Kendall* τ correlation is the following one:

$$\tau = \frac{n_c - n_d}{n\,(n-1)/2}\ , \quad -1 \leq \tau \geq 1 \tag{24}$$

$n\,(n-1)/2$ corresponds to the total number of possible pairings of ***x1*** with ***x2*** observations; $\mathrm{n_c}$ is the number of concordant pairs, i.e., if both members of one observation are larger than their respective members of the other observations. In other words, any pair of observations (***x1*** and ***x2***) are concordant if both $x_{1,x} > x_{2,x}$ and $x_{1,y} > x_{2,y}$ or $x_{1,x} < x_{2,x}$ and

$x_{1,y} < x_{2,y}$. Otherwise the pairs are discordant, if the two numbers in one observation differ in opposite directions (n_d).

Thus, in this case, the synchronization index (τ) provides information about the correlation between pairs of markers in order to infer, for example, the contraction timing in motion (if both coordinate values synchronously increase or decrease).

The information from both symmetry and synchronization index are of utmost importance since the clinician can quantitatively evaluate static and dynamic features with a single descriptive value and in latter case determine which face side, muscle group or component (horizontal or vertical) may suffer an intervention during a subsequent plastic surgery, for example.

4.2 Results and Discussion

Once obtained the *facegram*s from both controls and patients, the results were empirically analyzed and compared. As way of examples, Appendix D illustrates the *facegrams* of control 1 and patient 1. However, the discussion of the results will focus on both patients and control 1. It should be noted that the content of the proposed *facegram* was approved by specialists in plastic reconstructive surgery, but improvements regarding the design and structure can be made (more condensed information, real size graphics and displayed in graph paper, etc.).

4.2.1 Control subject 1

Through visual observation, there are no distinguishable features of asymmetry, either at rest or fully extension (see Figure D1 in Appendix D, initial set of images). Apparently, all points are symmetrical in relation to facial axis, with great amplitude of movements for commissures and midpoints.

Regarding static analysis, both tracings show that the cupid was coincident with facial symmetry axis (*y*-axis) and the relative position between landmarks was approximately the same for both motion stages. At rest, the contralateral points of each pair (commissures and midpoints) had a deviation lower than 3 mm (for both *x* and *y* coordinates). In turn, at maximum extension frame, the differences between the positions of contralateral points increased. Still, the symmetry index obtained for this subject was close to 1 for both pairs. The maximum extension results (first table in Figure D1, Appendix D) show that, as expected, the anatomical points that suffered greater extent were the commissures landmarks. In addition, the amplitude of excursion is relatively greater at the left side, with differences that do not exceed 8mm.

Focusing on dynamic analysis, the graphical tracings that correspond to horizontal, vertical and depth displacement as function of time show, in general, coordination between movement of facial points, with relaxation and contraction stages overlapped and apparent temporal symmetry. The movement has wider amplitude on vertical component and is consistent over time. In *Rest* and *Hold* stages, the point's positions are properly maintained, with larger fluctuations in the depth component. Regarding horizontal displacement (first plot), it is possible to observe that, at the initial point, contralateral markers are relatively symmetric in relation to symmetry axis, as well as the cupid is, once again, coincident with the same. The vertical displacement plot shows that, at the first instant, the contralateral markers were distanced about 2 mm from each other and this distance decreases during

Contraction, *Hold* and *Relaxation* phases. Finally, the evolution of depth displacement was consistent over time, with more fluctuations, and, at rest, commissures were naturally furthest from the cupid that midpoints, with differences smaller than 5 mm between contralateral markers of each pair. Through graphic analysis, it can be concluded that, once more, the movement was more pronounced at commissures, for all components.

The last two pair of graphs relate to parametric trajectories constructed for each pair of markers (commissures and midpoints), showing each individually excursion. As stated in section 3.2.6, in a normal excursion, the trajectory must be represented by a smooth line through time, without abrupt changes in direction and preferably beginning and finishing at the same place. In the case of control 1, and as shown in Figure D1, in Appendix D, the path traveled by commissures points was relatively smooth, with few abrupt changes and slight displacement between initial and final points. The right commissure trajectory showed marginally different behaviors in contraction and relaxation phases, as opposed to the contralateral, confirming slight asymmetry in these stages. It is also possible to observe that the markers had an insignificant asymmetry at rest, as confirmed by previous results. The situation is similar regarding the midpoints pair: right midpoint with slight asymmetry in contraction and relaxation phases and, for both, initial and final points had different positions.

By analyzing the velocity profile graph, it is possible to verify the presence of oscillations in the first section of markers' trajectory. In this phase, theoretically, the velocity should be approximately zero for a standard subject. However, these results can be interpreted taking into account two different aspects: the subject may have failed to maintain absolute rest position, subtly moving his mouth during this phase and, allied to this, the presence of local fluctuations in the tracked positions of the markers, over time, may have caused these instabilities. In general, it appears that marker's velocity increases in *Contraction* and *Relaxation* phases, as expected. Besides that, in *Rest* and *Hold* phases, the velocity is close to zero, with fluctuations that do not exceed 4 mm/s, in average.

Finally, in terms of synchrony between each pair of markers in each component (horizontal and vertical), the results show that the degree of dependence among markers is high (*Kendall's tau* close to 1) for vertical movement and for both pair 1 to pair 2. This means that the pairs of markers are synchronized along vertical movement. As regards the horizontal component (*x*), the obtained values are closer to -1 and, therefore, a negative correlation is observed, still lower to the vertical component (in absolute value).

4.2.2 Patient 1

In this patient, and merely by visual analysis, it is possible to detect some asymmetry at rest, especially among the commissures. In addition, during the movement the left side of the face has, apparently, greater amplitude of excursion motion, with signs of exaggerated facial contraction.

Using the *facegram*, besides being possible to observe the position of different anatomical points in these stages, it is possible to quantify their movement, which appears to be asymmetric and not synchronized.

In the resting phase, of which the first tracing of static analysis is an example, a cupid deviation to the left in relation to symmetry axis (y-axis) is observed (see Figure D2, Appendix D). The midpoints were aligned with the cupid but not with the facial symmetry axis; however, commissures had horizontal symmetry at rest and low vertical asymmetry. The

same applies to maximum extension instant, captured in the second graphic: midpoints asymmetry relative to facial axis and cupid deviation; slight horizontal and vertically asymmetry of the commissures. Thus, results using visual observation of symmetry at rest are refuted by static analysis graphics shown in *facegram*. On the other hand, the amplitude of excursion was greater at the left side, as shown in the first table of Figure D2, supporting the conclusions extracted by visual observation. As expected, the results of maximum extensions for each anatomical landmark were lower when compared to values obtained for control 1, standard subject. There is some symmetry of the commissures at maximum extension instant, confirmed by the obtained index of symmetry, which was relatively close to 1. Rather, there is no symmetry of midpoints at maximum extension, once right midpoint is closer to the facial axis when compared to left midpoint.

Dynamic analysis allowed the comparison between vertical, horizontal and depth displacement over time, in order to detect temporal consistency between each anatomical point and between spatial components. The point's movement was almost exclusively vertical for all markers, with some fluctuations in the horizontal component, and the commissures also had quite extensive depth of movement. As mentioned above and also concluded from these tracings, there is horizontal symmetry, but slight vertical asymmetry, of commissures at rest. Cupid has a deviation and there is horizontal asymmetry of midpoints at rest positions. From the analysis of all displacement's graphics, it is possible to conclude that initial and final positions of anatomical points were not the same, which may suggest, as stated in section 3.3.6, the presence of remaining contraction or partial relaxation in the last movement phase.

Regarding vertical displacement, both pairs [midpoints and commissures] appear to have synchronization over time and the right commissure started at a lower position than the contralateral marker, showing vertical asymmetry at rest, as mentioned. Besides that, during the *Hold* period, the vertical displacement was not conserved, with loss of extension, which prolongs until *Rest* phase.

The dynamic analysis can be complemented with the paths followed by each anatomical landmark. The pathways of the first pair, commissures, are described by non-smooth lines, with abrupt changes of direction and whose initial and final points do not match. The movement is not consistent over time and there is some disparity between *Contraction* and *Relaxation* stages. The conclusions are the same for the second pair of markers, whose lines are even less smooth, with many variations and smaller amplitude of excursion. In fact, the coordination can be quantified using a correlation coefficient, which for the case of horizontal movements, is close to 0 (unlike the standard subject), showing sparse dependence between studied variables. However, both pairs of markers showed good vertical movement coordination, with synchronization index close to 1, still lower than those obtained for the standard subject. This information is of extremely clinical value for the physician since it can provide information about the need to compensate the horizontal component during the next potential surgery, rather than vertical component.

In this case, the velocity profile of each marker is considerably more diffuse when compared to the standard subject. The initial *Rest* phase was properly maintained, with less fluctuations than control 1. Again, the highest velocities were found for *Contraction* phase, specifically for the right commissure marker. In fact, in *Contraction* and *Relaxation* phases, the speed of the right commissure is always superior to other markers, suggesting that the movement was uncoordinated and faster for the right side of the face, even if the maximum

extents have been obtained for the left side. In the final *Rest* phase, the velocity was not kept constant or near zero, which may indicate, once again, the presence of movement that is not expected at this stage.

4.2.3 Patient 2

By visual analysis of the facial movement of patient 2, during smile 2, it is possible to ascertain the presence of some disparity at rest, although not very pronounced, between each pair of markers. In addition, the amplitude of excursion seemed to be greater at the left side of patient's face and almost no movement at the right side. In fact, the movement caused by contraction apparently was initiated first for left markers and therefore the patient appeared to have lack of synchronization between the anatomical points of each pair. In this case, the phenomenon of synkinesis was visible: during smiling, the eye muscles involuntary contracted, leading to eye closure.

When analyzing the graphs that correspond to static analysis in *facegram* (data not shown), a slight asymmetry at rest is indeed observed, with differences below 6 mm for the horizontal and vertical components, when comparing both markers of each pair (commissures and midpoints). Furthermore, cupid has a deviation to the right of approximately 4 mm, in relation to symmetry axis. At maximum extension, the differences between the positions of each anatomical point in a pair are lower, with a shorter cupid's deviation in relation to face axis. In fact, in the graphical representation of maximal excursion (data not shown), the right commissure and right midpoint have the same positions as at rest. This can be explained if the frame captured in this tracing is prior to the initiation of contraction for the right side of the face, which possible indicates lack of temporal synchronization for both face sides, as observed visually. The weak synchronization will be later confirmed, during dynamic analysis. Quantitatively, the low amplitude of motion is supported by maximum extension values, being less than 1 cm for any marker. Furthermore, the excursion of the right side markers is actually lower than the ones at the left side, and even lower than the cupid movement. In general, the amount of movement is considerably lower than the one obtained for a standard subject. Nevertheless, the results from symmetry index show that, during the maximum extension, both commissural points and midpoints do not present symmetry, with symmetry index values near 0.5.

According to the obtained results for vertical, horizontal and depth displacement over time, facial motion is almost exclusively vertical, especially on the left side of the face. The horizontal movements are abrupt and with no apparently coordination. In all tracings, a spike is visible, during the contraction phase, which may indicate possible movement of the mouth or, if it was not the case, the presence of a muscle spasm. Merely by visual analysis it was impossible to verify this phenomenon. Besides this, the commissures and right midpoint were able to maintain the position during *Hold* phase, with possible spasmodic incident on the right side. For midpoints and commissures, some imbalance between *Contraction* and *Relaxation* phases is observed. The final positions of the various markers are different from the original ones, which may suggest, once again, partial relaxation in the final *Rest* phase.

The last graphical representation (data not shown), i.e., the pathway of each marker, grouped by pairs, confirm the differences between *Contraction* and *Relaxation* stages. Additionally, the curves are far from being smooth, with too many abrupt changes. This can be explained, as mentioned above, by muscle spasm incidents or involuntary contraction of

mouth muscles (in this case, the *orbicularis oris*) which may lead to a combined change of anatomic points' position and cause the observed spikes in the graphical representations.

Once again, the velocity profile representation shows more oscillations when compared to the standard subject. In any case, the *Rest* phases (initial and final) and the *Hold* stage exhibit speed fluctuations that may indicate the absence of motion stabilization. In the case of *Hold* phase, the velocity profile may also confirm the presence of spasmodic episodes, as described above. Apart from that, and in general, the *Contraction* and *Relaxation* phases exhibit higher velocity values for all markers, specifically for the left commissural marker.

Finally, and as expected, synchronization indices were low, compared to the standard subject, with higher values for midpoints pair in the vertical component. In this particular case, since the results are unsatisfactory, the next surgical procedures should be considered in order to improve the movement coordination and increase the degree of freedom of the facial excursion.

4.3 Conclusion

The lack of quantitative methods for assessing static, dynamic and 3-dimensional facial movements encouraged the development of a set of methods to track the anatomical points over time. From the obtained results, it was possible to organize the information in a report, proposed to be called *facegram*, which intends to be a set of dynamic and static morphological measures and representations that, through 3D information, helps the physician in measuring and assessing the movement of anatomical facial points' constellations.

This tool is divided into two sections, static and dynamic analysis, each of which presents illustrative graphs either of point's positions at rest and at maximum extension as well as trajectories of the different anatomical markers over time. Furthermore, two quantitative measures, symmetry and synchronization indexes were created for assessing, respectively, the degree of symmetry and correlation/dependency (the latter, in both vertical and horizontal components), between pairs of markers (commissures and midpoints).

Examples of *facegram*s of standard subjects and patients with pathologies were obtained and analyzed, aiming to extract as much information as possible about their facial movements. In the case of standard subject, with no apparent asymmetries, the trajectories of the anatomical markers were consistent over time, with overlapping of contraction and relaxation phases and synchronization between the two pairs of markers. However, it revealed slight asymmetries of the midpoints, both at rest and maximum extension. Therefore it is possible to conclude that, using the proposed method, even a smile that was visually symmetrical, showed some asymmetries and variability, mainly in the horizontal component, at micro-scale. The patients' results showed significant differences compared to control subject with some asymmetries at rest and at maximum extension, low amplitude of excursion and inconsistent movements with imbalance between contraction and relaxation phases.

This new strategy to assess the facial movement and present the results, both static and dynamic (*facegram*), appears to be applicable for quantification of movement in a complete and adequate way, being easy to interpret and with medical usefulness. The *facegram*, in terms of content, is the result of exhaustive discussion with the clinicians of *Plastic Surgery, Reconstructive and Maxillofacial* center (SCPRMF), at *Hospital São João* (Porto), and has the

required skills to be a universal and standardized tool for measuring facial excursion, in the medical context, with a user-friendly interface. There is, however, space for improvements regarding the design and structure of *facegram*.

Chapter 5

Conclusion

Facial disfigurement is a condition that affects thousands of people all over the world, and the increase in the number of cases encourages the constant improvement of reconstructive plastic surgery techniques aiming to minimize both functional and social consequences on patients. In fact, the main goal of these techniques is to restore facial symmetry and spontaneous and natural smile, potentially lost in the context of disease/accident. Therefore, is of extreme importance to assess, quantitatively and precisely, the motion restrictions caused by the facial disfigurement, in order to evaluate the effectiveness of the surgical procedures and even to compare medical, physical and surgical therapies.

Nevertheless, nowadays, the widely used methods that allow the evaluation of the outcomes after surgical interventions and, thus, the comparison between different techniques are subjective and ambiguous, not universally accepted as a standard procedure. To overcome these limitations, some studies have attempted to develop methodologies that enable objectively and quantitative analysis of face excursion; however, its high computational cost, expensive equipment, lack of dynamic and/or 3D information and poor clinical usability, make it impractical in the medical field. Consequently, the medical community does not have universal and standardized tools to measure the movement of the patients' face, before and after surgery, in a complete and accurate way and with a user-friendly interface.

Having this in mind, the main goal of this research was the study and development of a system capable of quantifying facial movement in a complete and accurate way as well as the creation of a novel strategy to present the clinical useful data. The first step was the development of a medical application for the acquisition of RGB and depth images, using a *Kinect*-based camera. Subjects, with and without apparent pathology, were recorded during the facial movements, in this case during the smile process. In the present work, physical markers were placed on the face of all subjects to identify the interest anatomical landmarks whose movement would be characterized. Thereafter, with the RGB and depth information, a set of methods was developed to perform tracking of points over time, using the *Block Matching Algorithm*. Techniques such as color distance transformation and marker enhancement pre-processing algorithms, as well as correction of head's movement have also been implemented. The spatial and temporal precision of the various components

(horizontal, vertical and depth) were tested. All methods used, as well as their variants, were tested regarding their capacity of efficiently perform tracking, i.e., through performance measures such as the coefficient of determination and the mean error. These measurements were carried out in comparison to theoretical trajectories obtained manually, from controls data. These validation tests allowed choosing the best set of methods to perform anatomical landmarks tracking. After determining the coordinates of interest points along frames, preliminary graphic representations of marker's trajectories were constructed and compared between patients and controls. The performance results prove that the proposed methodology is efficient in tracking the anatomical landmarks over time, allowing the creation of an efficient system to quantify facial movement.

The other aim of this research was the creation of a *facegram*, which can be easily interpreted by physicians and provide useful information for facial movements' quantification, either through static (symmetry at rest or maximum length) and dynamic (degree of synchronization) analysis. In fact, the challenge of this last part of the research was to create and choose which were the most appropriate measures and with clinical utility that should be presented in *facegram*. The content of this report was approved by specialists in plastic and reconstructive surgery, upon a statement drawn up by them (in Appendix E). Along with that, the author believes it has the skills needed to become a universal and standard tool to quantitatively measure facial movements. Some improvements can be done regarding the design and structure of facegram.

In the Literature Review section, various published quantitative methods were presented, as well as its limitations. With the proposed system to track facial points and analyze their movements in a quantitatively way, it was intended to overcome these limitations and develop an alternative solution, more appropriate to achieve these objectives.

Therefore, the suggested method presents advantages over others mentioned in section 2. The first one is the use of a low cost and portable acquisition system, such as Kinect-based cameras (in this case, *Asus Xtion Pro Live* camera). Thus, the whole acquisition structure can be integrated in a medical environment in an economically viable way, without requiring expensive equipment and without requiring specific installation setups, unlike other available methods [11, 26, 34]. These cameras have a market price on the order of 150 EUR.

Another feature, perhaps the most important one, is the fact that the proposed method operates in a 3D environment, i.e., collects and processes information from both intensity and depth, and the results are complemented with relevant 3D information. Besides that, it provides dynamic information about points' movement, using graphic representations of points' trajectories over time, as well as other useful medical information (such as symmetry and synchronization indexes, between pairs of points). All useful information is published on a report by the click of a button and whose content was thoroughly discussed and validated by specialists. These are advantages that make the presented method superior to its competitors [21, 22, 24, 25, 27, 28, 30, 32, 53, 54, 56].

A drawback that many previously described methods presented was the fact that the implemented algorithms were time consuming [1, 22, 27, 28, 30-33]. In the present case, the whole algorithm containing all methods used to track points over several frames, including RGB and depth pre-processing, as well as the correction of head movement, has a relatively low computational cost, taking approximately 2 minutes to analyze a set of 250 frames (~10

seconds of video recording), in a computer with the following hardware specifications: *Windows 7*; *64 bit(x64)* processor; *Intel® Core™ i5-3317U CPU@ 1.70 GHz* and *4GB RAM*.

The fact that the proposed system includes a kinematic model, allows further decrease in computational time, in cases in which facial movements are of great amplitude between frames. In the images analyzed in the present research, the kinematic model showed no significant influence regarding the computational time, since the location changes of the anatomical points are of subpixel (as a result of a combination of slow markers movement regarding the acquisition frame rate, ~25Hz, and low image/spatial resolution – 640x480px) for the used camera.

Additionally, as mentioned, the undesirable movement of the head of the patient can be corrected using a proposed method, which makes it more robust and useful on real-environment clinical situations.

Finally, the code behind the presented method was constructed to be as versatile as possible, allowing:

- The selection of any other constellations of physical markers, with no obligation to be the mouth anatomical points;
- The selection of any points, without the need to pre-defined physical markers on patient's face.
- The acquisition of RGB and depth images using any other type of depth camera, that allow different resolution/bit-depth/frame rate.
- Multiregional analysis (point's constellations), unlike some methods available on the market.

Despite overcoming most of the strong limitations presented for the available methods of quantification of facial movement, the system proposed in this research has also, of course, some drawbacks that should be mentioned.

The first is the fact that it is not a fully automatic method. The user intervention is necessary to initially choose one or more points intended to be tracked. For this reason, it is important to bear in mind that the manual selection of points, essential in this case, is operator-dependent and hard to control and, then, is the main source of errors. This is true because the choice of initial points is critical in the quality of tracking, since the neighborhood's and point's information are used to perform tasks like RGB pre-processing and tracking itself.

Another major limitation found in the course of this research was the low camera resolution, which hinders the tracking of points and especially the pre-processing, affecting final results.

In addition, as the *Block Matching Algorithm* depends only on RGB and depth information from template and search blocks, obtained after pre-processing, and since the intensity values vary from frame to frame, tracking errors as the ones that occurred for some set of images in control 1 are likely to happen, i.e., tracking deviation for more similar areas to the template block but different from the desired location. In order to overcome this, one solution is to increase camera resolution, as mentioned, in order to reduce intensity fluctuations. Another one is to transform the BMA in order to become more robust and incorporate other features, useful for tracking. One example is to combine the template

block with information about the first frame, as is already done, but also with previous frame information, to prevent errors.

Table VI gives an overview of the presented advantages and limitations of the proposed system, in comparison to the most relevant and published methods, regarding features that are essential to characterize the complex facial movements in a clinical environment.

Table VI. Comparative analysis between the more relevant methods for quantifying facial movement, and the system proposed in this research [1, 11, 21, 22, 26, 28, 34, 56].

	Dynamic analysis	3D analysis	Low cost of equipment	High spatial resolution (< 1mm)	Easy- to-use	Low Computational cost	Adapted for clinical environment
Frey et. al., 1999	X	✓	X	X	X	NA	X
Wachtmann et. al., (2001)	✓	X	NA	NA	✓	NA	X
Linstrom (2002)	✓	X	NA	NA	✓	X	X
Tomat and Manktelow (2005)	X	X	✓	X	✓	X	X
Hontanilla et. al. (2008)	✓	✓	X	✓	✓	✓	X
Mehta et. al. (2008)	✓	✓	X	✓	X	NA	X
Hadlock et. al. (2012)	X	X	NA	NA	✓	NA	X
Tzou et. al. (2012)	✓	✓	NA	NA	X	X	X
Proposed system	✓	✓	✓	X	✓	✓	✓

NA: not available information.

Nevertheless, taking into consideration the performance results and the outcomes with clinical usefulness from the *facegram* and the advantages over other methods, the proposed system can be a solution for tracking and quantification of facial motion anatomical landmarks, which likely to be integrated in a medical environment.

5.1 Future work

Although the results obtained after tracking are very promising and the measurements listed in *facegram* are capable of being translated into useful medical information, there are some improvements that could be identified, as well as suggestions of research topics that can be exploited with the obtained results.

The first one concerns the proposed methodology for color-map contrast enhancement (pre-processing), used to increase the local contrast of the markers. In the proposed system, a new image is obtained in which the value of each pixel is the distance to the RGB color space between the original values and the R, G and B reference, obtained from the central point and neighborhood points of each marker. A new approach to this pre-processing would be calculating the distance, not in relation to RGB color space, but to HSL color space. This color system is formed by hue, saturation and luminosity components and is a representation

that aims to be more intuitive and relevant than the Cartesian representation (RGB), rearranging the RGB geometry for the purpose. This system could have advantages in cases of low or high illumination of data acquisition spaces. This is because, with more or less light, the color variation in this system would affect only a specific component (luminosity), and could be easier to control the contrast in these cases. Despite the required initial conversion to HSL space, and again RGB conversion, where additional precision errors could be introduce in the system, it would be interesting to test the performance of this alternative and compare with the remaining used.

In order to improve the points' tracking, and to ensure method precision even for larger movements between frames it would be interesting to modify the proposed kinematic model to include the acceleration component and, thus implement a kinematic model of 2^{nd} order. Another option would be to implement *Kalman* filters.

Since the resolution of the depth camera used is critical for a precise tracking of facial anatomical points, the acquisition and analysis of frames using a camera with higher resolution, as is the case of *Microsoft Kinect v2,* would be substantially improved and better results would be obtained, with less fluctuations and noise.

Regarding the *facegram* representation, another feature that could be included in this system is the determination of synkinesis phenomenon. Although there is already a study in which one of the objectives is to detect these movements [28], the outcome of these measurements does not provide information of great medical utility. Therefore, it would be necessary to design a measure which allows the quantification of these particular involuntary movements, so important for facial asymmetry and coordination.

A different area of research that could be exploited using the results from the implemented tracking methods is the study of facial muscles contraction. The ultimate goal of this study would be the creation of a method that uses the positions of the multiple landmarks to infer the underlying state of contraction of the facial muscles. In fact, these results would further complement the present project, increasing its value in the medical field.

References

1. Tzou, C.-H.J., et al., *Evolution of the 3-Dimensional Video System for Facial Motion Analysis: Ten Years' Experiences and Recent Developments*. Annals of plastic surgery, 2012. **69**(2): p. 173-185.
2. *Changing Faces, Statistics*. 2014; Registered Charity No. 1011222]. Available from: https://www.changingfaces.org.uk/Health-Care-Professionals/Introduction-to-patient-needs/Statistics.
3. Hoffman, W.Y., *Reanimation of the paralyzed face*. Otolaryngol Clin North Am, 1992. **25**(3): p. 649-67.
4. Kosins, A.M., et al., *Facial paralysis for the plastic surgeon*. Can J Plast Surg, 2007. **15**(2): p. 77-82.
5. Horta, R., et al., *A facegram for spatial-temporal analysis of facial excursion: Applicability in the microsurgical reanimation of long-standing paralysis and pretransplantation*. J Craniomaxillofac Surg, 2014.
6. Levine, E., et al., *Quality of life and facial trauma: psychological and body image effects*. Ann Plast Surg, 2005. **54**(5): p. 502-10.
7. Horta, R., et al., *Facial disfigurement: restoration of facial dynamics in a patient with concomitant facial paralysis and blindness*. J Reconstr Microsurg, 2014. **30**(3): p. 207-10.
8. Moss, T.P., *The relationships between objective and subjective ratings of disfigurement severity, and psychological adjustment*. Body Image, 2005. **2**(2): p. 151-9.
9. Hagedoorn, M. and E. Molleman, *Facial disfigurement in patients with head and neck cancer: the role of social self-efficacy*. Health Psychology, 2006. **25**(5): p. 643.
10. Papel, D.I., et. al., *Facial Plastic and Reconstructive Surgery*. 3rd ed. 2009, New York: Thieme Medical Publishers, Inc. .
11. Hontanilla, B. and C. Auba, *Automatic three-dimensional quantitative analysis for evaluation of facial movement*. J Plast Reconstr Aesthet Surg, 2008. **61**(1): p. 18-30.
12. Kosowski, T.R., et al., *A systematic review of patient-reported outcome measures after facial cosmetic surgery and/or nonsurgical facial rejuvenation*. Plastic and reconstructive surgery, 2009. **123**(6): p. 1819-1827.
13. Ross, B.G., G. Fradet, and J.M. Nedzelski, *Development of a sensitive clinical facial grading system*. Otolaryngol Head Neck Surg, 1996. **114**(3): p. 380-6.
14. House, J.W. and D.E. Brackmann, *Facial nerve grading system*. Otolaryngol Head Neck Surg, 1985. **93**(2): p. 146-7.
15. Alsarraf, R., et al., *Measuring cosmetic facial plastic surgery outcomes: A pilot study*. Archives of Facial Plastic Surgery, 2001. **3**(3): p. 198-201.
16. Ching, S., et al., *Measuring outcomes in aesthetic surgery: a comprehensive review of the literature*. Plastic and reconstructive surgery, 2003. **111**(1): p. 469-482.
17. Alsarraf, R., *Outcomes research in facial plastic surgery: a review and new directions*. Aesthetic plastic surgery, 2000. **24**(3): p. 192-197.

18. Coulson, S.E., et al., *Reliability of the "Sydney," "Sunnybrook," and "House Brackmann" facial grading systems to assess voluntary movement and synkinesis after facial nerve paralysis.* Otolaryngol Head Neck Surg, 2005. **132**(4): p. 543-9.
19. Ginat, D.T., et al., *Facial reanimation procedures depicted on radiologic imaging.* AJNR Am J Neuroradiol, 2014. **35**(9): p. 1662-6.
20. Samsudin, W.S.W. and K. Sundaraj, *Clinical and non-clinical initial assessment of facial nerve paralysis: A qualitative review.* Biocybernetics and Biomedical Engineering, 2014. **34**(2): p. 71-78.
21. Wachtman, G.S., et al., *Automated tracking of facial features in patients with facial neuromuscular dysfunction.* Plast Reconstr Surg, 2001. **107**(5): p. 1124-33.
22. Tomat, L.R. and R.T. Manktelow, *Evaluation of a new measurement tool for facial paralysis reconstruction.* Plast Reconstr Surg, 2005. **115**(3): p. 696-704.
23. Yuen, K., et al., *Evaluation of facial palsy by Moire topography*, in *The Facial Nerve*. 1994, Springer. p. 541-544.
24. Fields, M.J. and N.S. Peckitt, *Facial nerve function index: a clinical measurement of facial nerve activity in patients with facial nerve palsies.* Oral Surgery, Oral Medicine, Oral Pathology, 1990. **69**(6): p. 681-682.
25. Sargent, E.W., O.A. Fadhli, and R.S. Cohen, *Measurement of facial movement with computer software.* Arch Otolaryngol Head Neck Surg, 1998. **124**(3): p. 313-8.
26. Mehta, R.P., S. Zhang, and T.A. Hadlock, *Novel 3-D video for quantification of facial movement.* Otolaryngology--Head and Neck Surgery, 2008. **138**(4): p. 468-472.
27. Isono, M., et al., *An objective evaluation method for facial mimic motion.* Otolaryngol Head Neck Surg, 1996. **114**(1): p. 27-31.
28. Linstrom, C.J., *Objective facial motion analysis in patients with facial nerve dysfunction.* Laryngoscope, 2002. **112**(7 Pt 1): p. 1129-47.
29. Burres, S., *Objective grading of facial paralysis.* The Annals of otology, rhinology, and laryngology, 1985. **95**(3 Pt 1): p. 238-241.
30. Wood, D.A., et al., *Objective measurement of normal facial movement with video microscaling.* Am J Otol, 1994. **15**(1): p. 61-5.
31. Meier-Gallati, V., H. Scriba, and U. Fisch, *Objective scaling of facial nerve function based on area analysis (OSCAR).* Otolaryngology-Head and Neck Surgery, 1998. **118**(4): p. 545-550.
32. Paletz, J.L., R.T. Manktelow, and R. Chaban, *The shape of a normal smile: implications for facial paralysis reconstruction.* Plast Reconstr Surg, 1994. **93**(4): p. 784-9; discussion 790-1.
33. Johnson, P.C., et al., *Simultaneous quantitation of facial movements: the maximal static response assay of facial nerve function.* Annals of plastic surgery, 1994. **32**(2): p. 171-179.
34. Frey, M., et al., *Three-dimensional video analysis of facial movements: a new method to assess the quantity and quality of the smile.* Plastic and reconstructive surgery, 1999. **104**(7): p. 2032-2039.
35. Zhang, S. and S.-T. Yau, *High-resolution, real-time 3D absolute coordinate measurement based on a phase-shifting method.* Optics Express, 2006. **14**(7): p. 2644-2649.
36. *United Network for Organ Sharing - UNOS.* 2014; Available from: http://www.unos.org/index.php.
37. Wayne F. Larrabee, K.H.M., Jenifer L. Henderson, *Surgical Anatomy of the Face*. 2004: Lippincott Williams & Wilkins.
38. Vicki Bruce, A.W.Y., *Face Perception*. 2012: Psychology Press. 481.
39. Prendergast, P.M., *Anatomy of the Face and Neck*, in *Cosmetic Surgery*. 2013, Springer. p. 29-45.
40. Netter, F.H., *Atlas of human anatomy*. 2010: Elsevier Health Sciences.
41. Siemionow, M., B.B. Gharb, and A. Rampazzo, *The face as a sensory organ.* Plast Reconstr Surg, 2011. **127**(2): p. 652-62.
42. Gray, H., *Anatomy of the human body*. 1918, Philadelphia: Lea & Febiger. p. 1396.
43. Rumsey, N. and D. Harcourt, *Body image and disfigurement: issues and interventions.* Body Image, 2004. **1**(1): p. 83-97.

44. Jones, B.C., et al., *Facial symmetry and judgements of apparent health: Support for a "good genes" explanation of the attractiveness–symmetry relationship.* Evolution and human behavior, 2001. **22**(6): p. 417-429.
45. Fink, B., et al., *Facial symmetry and judgements of attractiveness, health and personality.* Personality and Individual Differences, 2006. **41**(3): p. 491-499.
46. Grammer, K. and R. Thornhill, *Human (< em> Homo sapiens) facial attractiveness and sexual selection: The role of symmetry and averageness.* Journal of comparative psychology, 1994. **108**(3): p. 233.
47. Wong, T.-Y., et al., *Comparison of 2 methods of making surgical models for correction of facial asymmetry.* Journal of oral and maxillofacial surgery, 2005. **63**(2): p. 200-208.
48. *The American Board of Cosmetic Surgery, Cosmetic Surgery vs. Plastic Surgery*. 2009-2014; Available from: http://www.americanboardcosmeticsurgery.org/patient-resources/cosmetic-surgery-vs-plastic-surgery/.
49. The Johns Hopkins University, T.J.H.H., and Johns Hopkins Health System. *Facial Plastic and Reconstructive Surgery at Johns Hopkins*. 2014; Available from: http://www.hopkinsmedicine.org/facial_plastic_reconstructive_surgery/reconstructive_procedures/facial_reanimation_surgery.html.
50. *American Academy of Facial Plastic and Reconstructive Surgery (AAFPRS), Facial Reconstructive Procedures*. 2015; Available from: http://www.aafprs.org/patient/procedures/facial_reconst.html.
51. *University of Maryland Medical Center, Facial paralysis.* 2014.
52. Samsudin, W.S.W. and K. Sundaraj, *Evaluation and Grading Systems of Facial Paralysis for Facial Rehabilitation.* Journal of Physical Therapy Science, 2013. **25**(4): p. 515-519.
53. Frey, M., et al., *Development of a new paresis scoring system for pre- and postoperative evaluation of facial paresis*. Eur Arch Otorhinolaryngol, 1994: p. S182-4.
54. Neely, J.G., et al., *Computerized quantitative dynamic analysis of facial motion in the paralyzed and synkinetic face*. Otology & Neurotology, 1992. **13**(2): p. 97-107.
55. Neely, J.G., et al., *Quantitative assessment of the variation within grades of facial paralysis.* The Laryngoscope, 1996. **106**(4): p. 438-442.
56. Hadlock, T.A. and L.S. Urban, *Toward a universal, automated facial measurement tool in facial reanimation.* Archives of facial plastic surgery, 2012. **14**(4): p. 277-282.
57. Bray, D., et al., *Assessing outcomes in facial reanimation: evaluation and validation of the SMILE system for measuring lip excursion during smiling.* Arch Facial Plast Surg, 2010. **12**(5): p. 352-4.
58. Popat, H., et al., *Quantitative analysis of facial movement—A review of three-dimensional imaging techniques.* Computerized Medical Imaging and Graphics, 2009. **33**(5): p. 377-383.
59. Vezzetti, E. and F. Marcolin, *3D human face description: landmarks measures and geometrical features.* Image and Vision Computing, 2012. **30**(10): p. 698-712.
60. Celiktutan, O., S. Ulukaya, and B. Sankur, *A comparative study of face landmarking techniques.* Eurasip Journal on Image and Video Processing, 2013.
61. Gupta, S., et al. *3D face recognition founded on the structural diversity of human faces*. in *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*. 2007. IEEE.
62. Gupta, S., M.K. Markey, and A.C. Bovik, *Anthropometric 3D face recognition.* International Journal of Computer Vision, 2010. **90**(3): p. 331-349.
63. Carnický, J. and D. Chorvát Jr, *Three-dimensional measurement of human face with structured-light illumination.* Measurement science review, 2006. **6**(1): p. 1.
64. Farkas, L.G., *Anthropometry of the Head and Face*. Raven Press, New York, 1994.
65. Farkas, L.G., *Accuracy of anthropometric measurements: past, present, and future.* The Cleft palate-craniofacial journal, 1996. **33**(1): p. 10-22.
66. Tie, Y. and L. Guan, *Automatic landmark point detection and tracking for human facial expressions.* EURASIP Journal on Image and Video Processing, 2013. **2013**(1): p. 1-15.

67. Li, Y., et al., *Simultaneous facial feature tracking and facial expression recognition*. Image Processing, IEEE Transactions on, 2013. **22**(7): p. 2559-2573.
68. Valente, S. and J.-L. Dugelay, *A visual analysis/synthesis feedback loop for accurate face tracking*. Signal Processing: Image Communication, 2001. **16**(6): p. 585-608.
69. Senechal, T., V. Rapp, and L. Prevost. *Facial feature tracking for emotional dynamic analysis*. in *Advanced Concepts for Intelligent Vision Systems*. 2011. Springer.
70. Cai, Q., et al., *3D Deformable Face Tracking with a Commodity Depth Camera*. Computer Vision-Eccv 2010, Pt Iii, 2010. **6313**: p. 229-242.
71. Yang, F., et al., *Robust Face Tracking with a Consumer Depth Camera*. 2012 Ieee International Conference on Image Processing (Icip 2012), 2012: p. 561-564.
72. Zhang, W., Q. Wang, and X. Tang, *Real time feature based 3-d deformable face tracking*, in *Computer Vision-ECCV 2008*. 2008, Springer. p. 720-732.
73. Gao, X., et al., *A review of active appearance models*. Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, 2010. **40**(2): p. 145-158.
74. Ong, E.-J. and R. Bowden, *Robust facial feature tracking using shape-constrained multiresolution-selected linear predictors*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2011. **33**(9): p. 1844-1859.
75. Cootes, T.F., et al., *Active shape models-their training and application*. Computer vision and image understanding, 1995. **61**(1): p. 38-59.
76. Cootes, T.F., G.J. Edwards, and C.J. Taylor, *Active appearance models*. IEEE Transactions on pattern analysis and machine intelligence, 2001. **23**(6): p. 681-685.
77. Baltrusaitis, T., P. Robinson, and L.P. Morency, *3D Constrained Local Model for Rigid and Non-Rigid Facial Tracking*. 2012 Ieee Conference on Computer Vision and Pattern Recognition (Cvpr), 2012: p. 2610-2617.
78. Sonka, M., V. Hlavac, and R. Boyle, *Image processing, analysis, and machine vision*. 2014: Cengage Learning.
79. Tomasi, C. and T. Kanade, *Detection and tracking of point features*. 1991: School of Computer Science, Carnegie Mellon Univ. Pittsburgh.
80. Jokinen, O., *Tracking of facial deformations in multi-image sequences with elimination of rigid motion of the head*. ISPRS Journal of Photogrammetry and Remote Sensing, 2013. **84**: p. 52-68.
81. Ter Braak, C.J., *A Markov Chain Monte Carlo version of the genetic algorithm Differential Evolution: easy Bayesian computing for real parameter spaces*. Statistics and Computing, 2006. **16**(3): p. 239-249.
82. Xiong, X. and F. De la Torre. *Supervised descent method and its applications to face alignment*. in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*. 2013. IEEE.
83. Sobottka, K. and I. Pitas. *A fully automatic approach to facial feature detection and tracking*. in *Audio-and Video-based Biometric Person Authentication*. 1997. Springer.
84. Sobottka, K. and I. Pitas, *A novel method for automatic face segmentation, facial feature extraction and tracking*. Signal processing: Image communication, 1998. **12**(3): p. 263-281.
85. Sengijpta, S.K., *Fundamentals of statistical signal processing: Estimation theory*. Technometrics, 1995. **37**(4): p. 465-466.
86. Koyuncu, B. and U.F. Akdogan, *Reconstruction of facial images by using facial motion and image processing techniques*. Journal of X-Ray Science and Technology, 2008. **16**(3): p. 159-169.
87. Kim, J.S., et al. *Facial feature tracking by robust face segmentation and scalable rotational BMA*. 2002.
88. Haker, M., et al. *Geometric invariants for facial feature tracking with 3D TOF cameras*. in *Signals, Circuits and Systems, 2007. ISSCS 2007. International Symposium on*. 2007. IEEE.
89. *An overview of range images by Helmut Cantzler; CVonline: The Evolving, Distributed, Non-Proprietary, On-Line Compendium of Computer Vision*

2014 Robert Fisher; Available from: http://homepages.inf.ed.ac.uk/rbf/CVonline/CVentry.htm.

90. Meers, S. and K. Ward, *Head-pose tracking with a time-of-flight camera*. Faculty of Informatics-Papers, 2008: p. 720.
91. Zhang, Z., *Microsoft kinect sensor and its effect*. MultiMedia, IEEE, 2012. **19**(2): p. 4-10.
92. Litomisky, K., *Consumer rgb-d cameras and their applications*. University of California, Riverside. Ano, 2012.
93. Vineetha, G. and S.C.L. Joseph, *Face Expression Detection Using Microsoft Kinect with the Help of Artificial Neural Network*. 2012.
94. Khoshelham, K. and S.O. Elberink, *Accuracy and resolution of kinect depth data for indoor mapping applications*. Sensors, 2012. **12**(2): p. 1437-1454.
95. Clark, R.A., et al., *Validity of the Microsoft Kinect for assessment of postural control*. Gait & posture, 2012. **36**(3): p. 372-377.
96. Li, B.Y., et al. *Using kinect for face recognition under varying poses, expressions, illumination and disguise*. in *Applications of Computer Vision (WACV), 2013 IEEE Workshop on*. 2013. IEEE.
97. Meers, S. and K. Ward. *Face recognition using a time-of-flight camera*. in *Computer Graphics, Imaging and Visualization, 2009. CGIV'09. Sixth International Conference on*. 2009. IEEE.
98. Ganapathi, V., et al. *Real time motion capture using a single time-of-flight camera*. in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. 2010. IEEE.
99. Gokturk, S.B. and C. Tomasi. *3D head tracking based on recognition and interpolation using a time-of-flight depth sensor*. in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*. 2004. IEEE.
100. Uddin, M.Z., *A depth video-based facial expression recognition system*. IETE Technical Review, 2012. **29**(2): p. 169-177.
101. Seddik, B., et al. *Unsupervised facial expressions recognition and avatar reconstruction from kinect*. in *Systems, Signals & Devices (SSD), 2013 10th International Multi-Conference on*. 2013. IEEE.
102. Zell, A. *Real time face tracking and pose estimation using an adaptive correlation filter for human-robot interaction*. in *Mobile Robots (ECMR), 2013 European Conference on*. 2013. IEEE.
103. MathWorks. *Intensity-Based Automatic Image Registration*. Available from: http://www.mathworks.com/help/images/intensity-based-automatic-image-registration.html.
104. Yang, N.-E., Y.-G. Kim, and R.-H. Park. *Depth hole filling using the depth distribution of neighboring regions of depth holes in the Kinect sensor*. in *Signal Processing, Communication and Computing (ICSPCC), 2012 IEEE International Conference on*. 2012. IEEE.
105. Arun, K.S., T.S. Huang, and S.D. Blostein, *Least-squares fitting of two 3-D point sets*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 1987(5): p. 698-700.

Appendix A

Evaluation of surgical interventions:
Quantitative methods

Table A1. Overview of existing quantitative methods to analyze facial movement. Adapted from [1, 11, 21, 22, 24-28, 30-34, 53, 54, 56].

Author	*Method / Acquired data*	*Disadvantages*
Fields and Peckitt, 1990	Measurement of differences between facial distances, at rest and movement. Facial nerve function index.	No dynamic information; lacks accuracy.
Neely et al., 1992	Digitized images, pixel-subtraction. 2D analysis of facial motion; dynamic strength-duration curve.	No measurement of direction of motion's magnitude.
Frey et. al., 1994	Tracking of specified facial landmarks at rest and maximal facial expressions, with *Faciometer*. Displacement of markers during motion.	No multiregional analysis neither dynamic motion.
Johnson et. al., 1994	Tracking of facial markers on photographs. Measurement of direction of markers' displacement.	Time consuming methods; absence of markers' trajectories between rest and maximum extension.
Paletz et. al., 1994	Tracking of facial markers around the mouth on image projections. Measurement of reference points' positions.	No dynamic information. Time consuming.
Wood et. al., 1994	Video micro-scaling. Quantification of movement between pre-defined lines.	No movement's velocity information; need for head's stability, high computation time.
Isono et. al., 1996	Tracking of facial markers on photographs. Measurement of direction of markers' displacement.	Absence of 3D information; large computation time.
Meier-Gallati et. al., 1998	Facial light reflectance. Facial-contour changes.	No measurement of direction of motion's magnitude; large computation time.

Sargent et. al., 1998	Digital photographs, tracking of facial markers, pixel-subtraction. 2D analysis of facial motion, ratio of affected areas	No 3D dynamic information; need for multiple user interventions.
Frey et. al., 1999	3D tracking of facial markers, using video. Measurement of 3D distances and movements.	Very expensive equipment.
Wachtmann et. al., 2001	2D video analysis, feature points marked on digital image. Speed, acceleration, direction of motion and displacement.	No 3D information.
Linstrom, 2002	Tracking of light-reflective facial markers. Quantification of markers' displacement, time pattern of motion.	Time consuming method; not fully automatic method; no 3D information.
Tomat and Manktelow, 2005	Video recording, tracing of facial markers frames. 2D analysis of facial movements, displacement and angles of motion.	Time consuming; no relevant and dynamic information.
Hontanilla et. al. 2008	3D tracking of reflective facial markers, using infrared-light cameras. 3D quantification of facial motion; multiregional analysis.	Very expensive equipment; complex setup.
Mehta et. al., 2008	3D shape determination using a video acquisition system. 3D quantification of facial movement.	Not fully automatic method; very expensive equipment.
Hadlock et. al., 2012	2D tracking of facial markers. Measurement of facial distances at rest and movement;	No dynamic information.
Tzou et. al., 2012	3D tracking of facial markers, using *Facialis* and *Facishow*. Graphic representation of markers' 3D trajectories.	Time consuming methods.

Appendix B

Protocol for RGB-D Image Acquisition
using a *Kinect*-based camera

This protocol includes information on the procedure and requirements to be adopted in order to acquire images for facial tracking points. All specifications concern the *Asus Xtion Pro Live* camera, a *Kinect*-based camera. However, the protocol can be adapted for other types of RGB-D camera.

1. Camera specifications

1.1 Hardware requirements

The computer to which the camera is connected should have the following requirements:

- *Windows 7, Windows 8, Windows XP, Windows Vista, Linux Ubuntu 10.10: X86*
- 32 bit or 64 bit processor
- Interface *USB* 2.0/ 3.0
- *OpenNI Software Development Kit* (SDK) bundled and *NITE*

1.2 Camera Limits

- Object Distance: 0.8m to 3.5m
- Horizontal viewing angle: 58°
- Vertical viewing angle: 45°
- Diagonal viewing angle: 70°

1.3 Acquisition Parameters

- RGB stream resolution: 640x480 pixels
- Depth stream resolution: 640x480 pixels
- Both RGB and depth information are captured at a maximum of 30 fps and at the same time
- Depth precision: 100μm (pixel format)
- RGB pixel format: RGB 888

2. Data collection

2.1 Local Environment

- The local should be well illuminated, so that the patient's face is clearly visible and the contrast between the markers and the face is maximum. Avoid shadows and different lighting zones on patient's face.
- The light should come from behind from the camera, perpendicular to the face of the patient.

2.2 Patient Positioning

- Camera should be mounted at least at 0.80 m from the patient.
- The patient may be sitting or standing up, as long as his face is centered in the camera plane and with the camera's optical axis perpendicular to the face plane.

2.3 Initial Parameters

Initially, the user must mark the points to be tracked, plus 2 reference points, on patient's face. In addition, the distance between these two should be measured, in millimeters.

2.4 Face Movements

It will be ask to the patient to smile, avoiding head movements, in a series of 5 phases with durations defined by the user: rest, contract, hold, relax and rest. These steps will always be guided by a graphical interface.

At the end, the data are automatically recorded into a *.mat* file, containing RGB and depth information, markers' distance, patient's name, RGB and depth time stamp and the duration of each phase. The file name is in the format: *PatientNameX.mat,* where *X* represents the number of consecutive recordings from the same patient.

Appendix C

Tracking of Facial Movements:
Validation of the algorithms

Table C1. R^2 and mean error (M.e.) results (first and second line for each method, respectively) for each marker, in control 1 and 2, regarding the color-map enhancement method. Ref.1 = Reference marker 1; Ref.2 = Reference marker 2; LC = left commissure. Values in bold represent the best result of each set for a given marker. Red Channel = filtering only the red channel, without additional processing; Median T=0 = median between the neighborhood points of all markers, with template size T equal to 0 (just the center point); Median T=1 = median between the neighborhood points of all markers, with template size T equal to 1; Median *LoG* T=1 = additional processing using Laplacian of Gaussian; Mean T=1 = mean between the neighborhood points of all markers, with template size T equal to 1.

		Marker 1 (Ref.1)		Marker 2 (Ref.2)		Marker 3 (RC)		Marker 4 (RMP)		Marker 5 (Cupide)		Marker 6 (LMP)		Marker 7 (LC)	
	Coord.	*x*	*y*	*x*	*y*	*x*	*y*	*x*	*y*	*x*	*y*	*x*	*y*	*x*	*y*
Red Channel	R^2	-3.81	**-0.70**	**-1.04**	0.31	0.68	**0.97**	0.67	-8.55	**1.01**	**0.98**	-5.49	-8.91	-10.56	-9.75
	M.e.	0.07	**0.06**	**0.06**	0.02	0.18	**0.05**	0.09	0.95	**0.07**	**0.04**	0.72	1.08	1.30	1.50

Control 1	Median T = 0	R^2	-4.18	-0.21	-1.41	0.37	**0.80**	0.95	0.50	-9.85	-44.65	-11.95	-5.38	-10.21	-10.39	0.89
		M.e.	0.06	0.04	0.06	0.02	**0.13**	0.08	0.11	0.99	0.40	1.01	0.70	1.17	1.15	0.13
	Median T = 1	R^2	-2.94	-1.21	-1.16	0.31	0.783	0.84	0.39	-7.96	-51.99	-11.14	-5.93	-10.43	-8.81	0.89
		M.e.	0.05	0.08	0.06	0.02	0.14	0.18	0.12	0.92	0.44	0.98	0.73	1.18	1.08	0.13
	Median LoG T=1	R^2	**-2.86**	-0.92	-1.39	**0.33**	-4.19	-5.64	0.16	**0.84**	-178.91	0.93	**0.94**	**0.87**	**0.85**	**0.92**
		M.e.	**0.05**	0.07	0.06	**0.02**	0.52	1.09	0.15	**0.14**	0.77	0.06	**0.05**	**0.14**	**0.17**	**0.11**
	Mean T=1	R^2	-3.73	-1.40	-1.24	0.31	-0.26	0.44	**0.76**	-1.46	-52.41	-10.62	-5.97	-10.29	-9.02	-8.52
		M.e.	0.06	0.08	0.06	0.02	0.29	0.29	**0.07**	0.46	0.44	0.96	0.73	1.17	1.14	1.37
Control 2	Red Channel	R^2	0.99	-0.45	0.99	0.99	0.99	0.92	0.99	**0.94**	0.99	**0.98**	0.99	0.98	0.99	0.99
		M.e.	0.27	0.21	0.12	0.09	0.09	0.08	0.07	**0.11**	0.06	**0.06**	0.07	0.10	0.13	0.08
	Median T = 0	R^2	0.99	-0.53	0.99	0.99	**0.99**	0.92	**0.99**	0.93	**0.99**	0.98	**0.99**	**0.98**	0.99	0.99
		M.e.	0.18	0.21	0.12	0.08	**0.08**	0.08	**0.07**	0.12	**0.06**	0.07	**0.06**	**0.09**	0.13	0.09
	Median T = 1	R^2	0.99	-0.53	0.99	0.99	0.99	0.90	0.99	0.93	0.99	0.98	0.99	0.98	0.99	**0.99**
		M.e.	0.22	0.21	0.11	0.08	0.10	0.10	0.08	0.11	0.07	0.06	0.07	0.10	0.14	**0.09**
	Median LoG T=1	R^2	**0.99**	**0.86**	**0.99**	**0.99**	0.99	**0.93**	0.99	0.92	0.99	0.98	0.97	0.95	**0.99**	0.99
		M.e.	**0.07**	**0.10**	**0.07**	**0.06**	0.15	**0.07**	0.07	0.12	0.09	0.06	0.34	0.13	**0.08**	0.10
	Mean T=1	R^2	0.99	-1.12	0.99	0.99	0.99	0.9	0.99	0.93	0.99	0.98	0.99	0.98	0.99	0.99
		M.e.	0.31	0.28	0.11	0.08	0.09	0.10	0.08	0.11	0.07	0.06	0.07	0.10	0.14	0.09

Table C2. R^2 and mean error (M.e.) results (first and second line for each set, respectively) for each marker, in control 1 and 2. Ref.1 = Reference marker 1; Ref.2 = Reference marker 2; LC = left commissure. Values in bold represent the best result of each set for a given marker.

			Marker 1 (Ref.1)		Marker 2 (Ref.2)		Marker 3 (RC)		Marker 4 (RMP)		Marker 5 (Cupide)		Marker 6 (LMP)		Marker 7 (LC)	
		Coord.	*x*	*y*	*x*	*y*	*x*	*y*	*x*	*y*	*x*	*y*	*x*	*y*	*x*	*y*
Control 1	Set 1	R^2	-3.063	-0.69	-0.97	0.23	**0.92**	0.91	-2.09	**0.87**	-198.07	0.93	0.89	**0.90**	0.64	0.88
		M.e.	0.06	**0.06**	0.05	0.02	**0.07**	0.14	0.32	**0.12**	0.81	0.06	0.08	**0.10**	0.24	0.14
	Set 2	R^2	-3.27	-0.83	**-0.72**	0.31	0.87	0.78	-3.062	-3.15	-129.87	0.92	-8.99	-414.9	0.63	0.859
		M.e.	0.06	0.07	0.05	0.02	0.11	0.169	0.32	0.56	0.57	**0.05**	0.65	7.20	0.24	0.15
	Set 3	R^2	-3.19	-0.81	-0.87	0.25	0.64	**0.93**	-0.35	-0.46	**-60.76**	-12.61	-8.38	-10.86	-6.40	-6.38
		M.e.	0.06	0.06	0.05	0.02	0.18	0.11	0.15	0.30	**0.48**	1.07	0.87	1.23	0.81	0.94
	Set 4	R^2	**-2.86**	-0.92	-1.39	**0.33**	-4.19	-5.64	**0.16**	0.84	-178.91	**0.93**	**0.94**	0.87	**0.85**	**0.92**
		M.e.	**0.05**	0.07	0.06	**0.02**	0.52	1.09	**0.15**	0.14	0.77	0.06	**0.05**	0.14	**0.17**	**0.11**
Control 2	Set 1	R^2	0.97	0.53	0.99	0.99	**0.99**	0.77	0.99	0.90	0.99	0.98	**0.99**	**0.96**	0.99	0.91
		M.e.	0.45	0.16	0.12	0.07	0.15	0.16	0.16	0.14	0.10	0.06	0.16	**0.11**	0.07	0.28
	Set 2	R^2	0.98	0.58	0.99	0.99	0.99	0.49	0.99	0.85	0.99	0.98	0.99	0.96	0.99	0.87
		M.e.	0.44	0.16	0.14	0.08	0.16	0.22	0.07	0.16	0.15	0.06	0.17	0.12	0.09	0.33
	Set 3	R^2	0.97	0.19	0.98	0.99	0.96	-4.13	0.99	0.87	0.99	0.98	0.99	0.96	0.38	-5.29
		M.e.	0.49	0.18	0.26	0.09	0.34	0.46	0.07	0.15	0.09	0.06	**0.16**	0.11	1.82	2.67
	Set 4	R^2	**0.99**	**0.86**	**0.99**	**0.99**	0.99	**0.93**	**0.99**	**0.92**	**0.99**	**0.98**	0.97	0.95	**0.99**	**0.99**
		M.e.	**0.07**	**0.10**	**0.07**	**0.06**	**0.15**	**0.07**	**0.07**	**0.12**	**0.09**	**0.06**	0.34	0.13	**0.08**	**0.10**

Table C2. R squared and mean error results (first and second line for each set, respectively) for each marker, in control 1 and 2. Ref.1 = Reference marker 1; Ref.2 = Reference marker 2; LC = left commissure. Values in bold represent the best result for a given marker, compared with Table C2 (last line for each control).

		Marker 1 (Ref.1)		**Marker 2 (Ref.2)**		**Marker 3 (RC)**		**Marker 4 (RMP)**		**Marker 5 (Cupide)**		**Marker 6 (LMP)**		**Marker 7 (LC)**	
	Coord.	***x***	***y***	***x***	***y***	***x***	***y***	***x***	***y***	***x***	***y***	***x***	***y***	***x***	***y***
Control 1	R^2	-2.28	-1.43	-0.49	-0.38	**0.49**	**0.71**	-58.18	-9.98	**-1.24**	0.59	0.73	0.76	0.28	0.74
	M.e.	0.05	0.09	**0.04**	0.04	**0.21**	**0.23**	1.37	0.96	**0.08**	0.17	0.13	0.16	0.36	0.22
Control 2	R^2	0.99	-2.64	0.89	0.95	**0.99**	0.54	0.87	-0.09	0.83	0.73	0.93	0.92	0.99	0.96
	M.e.	0.14	0.40	0.96	0.16	0.15	0.21	0.73	0.36	0.79	0.24	0.58	0.16	0.21	0.20

Table C4.R^2 and mean error (M.e.) results (first and second line for each set, respectively) for each marker, in control 1 and 2, regarding the minimization function used in BMA algorithm. Ref.1 = Reference marker 1; Ref.2 = Reference marker 2; LC = left commissure. Values in bold represent the best result of each set for a given marker. Minimum = minimum of errors' sum; Median = minimum of point's neighborhood median; Conv1 = minimum of error sum's convolution; Conv2 = minimum of the sum between depth error convolution and RGB error convolution.

			Marker 1 (Ref.1)		Marker 2 (Ref.2)		Marker 3 (RC)		Marker 4 (RMP)		Marker 5 (Cupide)		Marker 6 (LMP)		Marker 7 (LC)	
		Coord.	*x*	*y*	*x*	*y*	*x*	*y*	*x*	*y*	*x*	*y*	*x*	*y*	*x*	*y*
Control 1	Minimum	R^2	**-2.86**	**-0.92**	-1.39	0.33	-4.19	-5.64	0.16	**0.84**	-178.91	0.93	**0.94**	0.87	0.85	**0.92**
		M.e.	**0.05**	**0.07**	0.064	0.0192	0.52	1.09	0.15	**0.14**	0.77	0.06	**0.05**	0.14	0.17	**0.11**
	Median	R^2	-2.98	-1.51	**-0.15**	-0.44	-6.21	-5.96	**0.40**	0.63	-183.23	**0.96**	0.95	**0.92**	**0.92**	0.83
		M.e.	0.05	0.09	**0.03**	0.04	0.76	1.13	**0.14**	0.21	0.77	**0.03**	0.04	**0.09**	**0.13**	0.21
	Conv1	R^2	-3.60	-1.21	-1.18	0.33	**0.16**	**0.51**	-0.25	-0.93	**-67.23**	-13.02	-66.34	-16.96	-10.48	-10.12
		M.e.	0.06	0.08	0.06	0.02	**0.21**	**0.26**	0.16	0.37	**0.49**	1.08	1.95	1.48	1.29	1.55
	Conv2	R^2	-3.60	-1.21	-1.18	0.33	**0.16**	**0.51**	-0.25	-0.93	**-67.23**	-13.02	-66.34	-16.9	-10.48	-10.12
		M.e.	0.06	0.078	0.06	0.02	**0.21**	**0.26**	0.16	0.37	**0.49**	1.08	1.95	1.48	1.29	1.55
Control 2	Minimum	R^2	**0.99**	0.86	**0.99**	**0.99**	**0.99**	**0.93**	**0.99**	**0.92**	**0.998**	**0.98**	0.97	0.95	**0.99**	**0.99**
		M.e.	**0.07**	0.098	**0.07**	**0.06**	**0.15**	**0.07**	**0.07**	**0.12**	**0.09**	**0.06**	0.34	0.13	**0.08**	**0.10**
	Median	R^2	0.99	0.87	0.99	0.99	0.99	0.92	0.99	0.90	0.99	0.97	0.98	0.92	0.99	0.98
		M.e.	0.11	0.08	0.08	0.06	0.15	0.07	0.08	0.14	0.09	0.10	0.30	0.19	0.09	0.16
	Conv1	R^2	0.84	-87.78	0.99	0.99	0.99	0.55	0.99	0.94	0.99	0.98	0.99	0.96	0.91	-0.35
		M.e.	1.02	1.72	0.12	0.08	0.13	0.14	0.08	0.11	0.08	0.06	0.12	0.11	0.68	1.14
	Conv2	R^2	0.84	-87.78	0.99	0.99	0.99	0.55	0.99	0.94	0.99	0.98	0.99	0.96	0.91	-0.35
		M.e.	1.02	1.72	0.12	0.08	0.13	0.14	0.08	0.11	0.08	0.06	0.12	0.11	0.68	1.14

Appendix D

Facegram
Control 1 and Patient 1

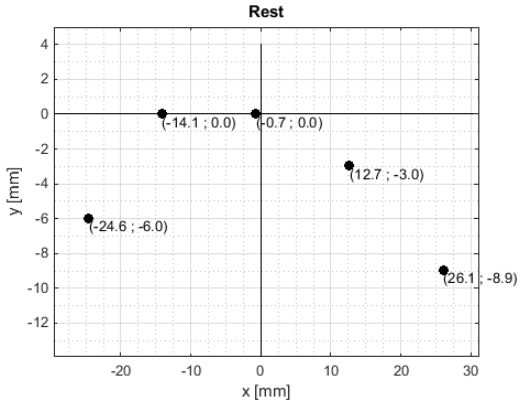
On the following pages will be illustrated two *facegrams* (control 1 and patient 1), in the same format with which they would be presented to the medical community. The *facegrams* will be named Figure D1 and Figure D2, respectively.

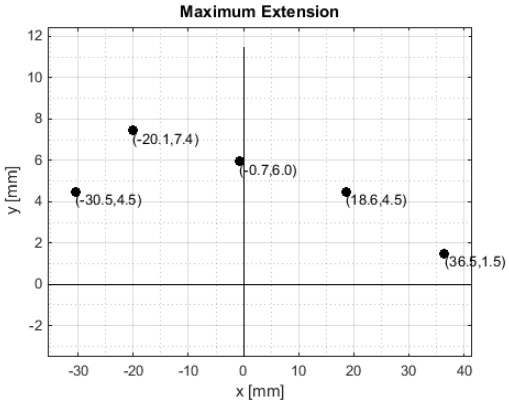
FACEGRAM.

Name: AnaFilipaTeste2

Sex: Female Age: 22 Date: 27-Jun-2015

Static analysis



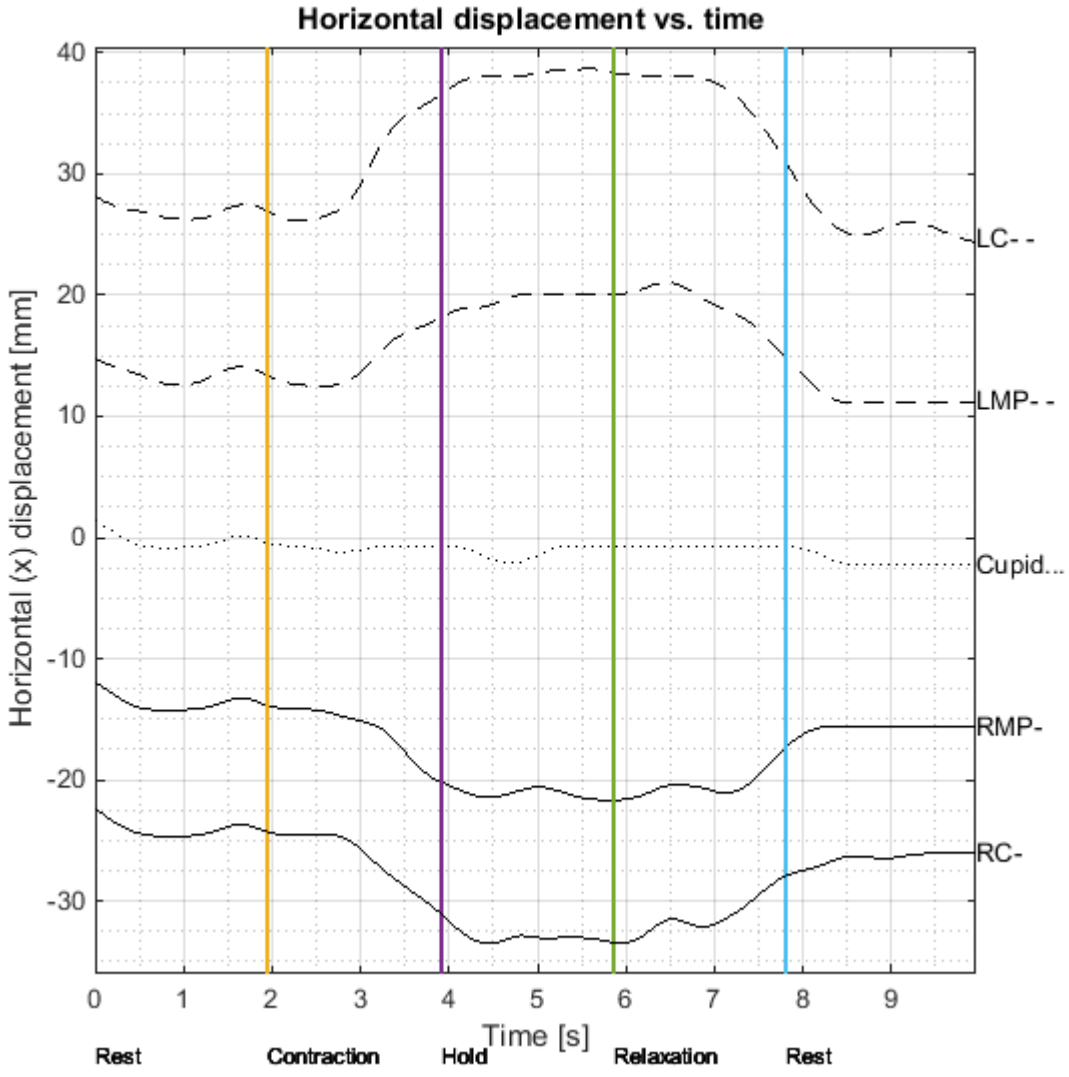


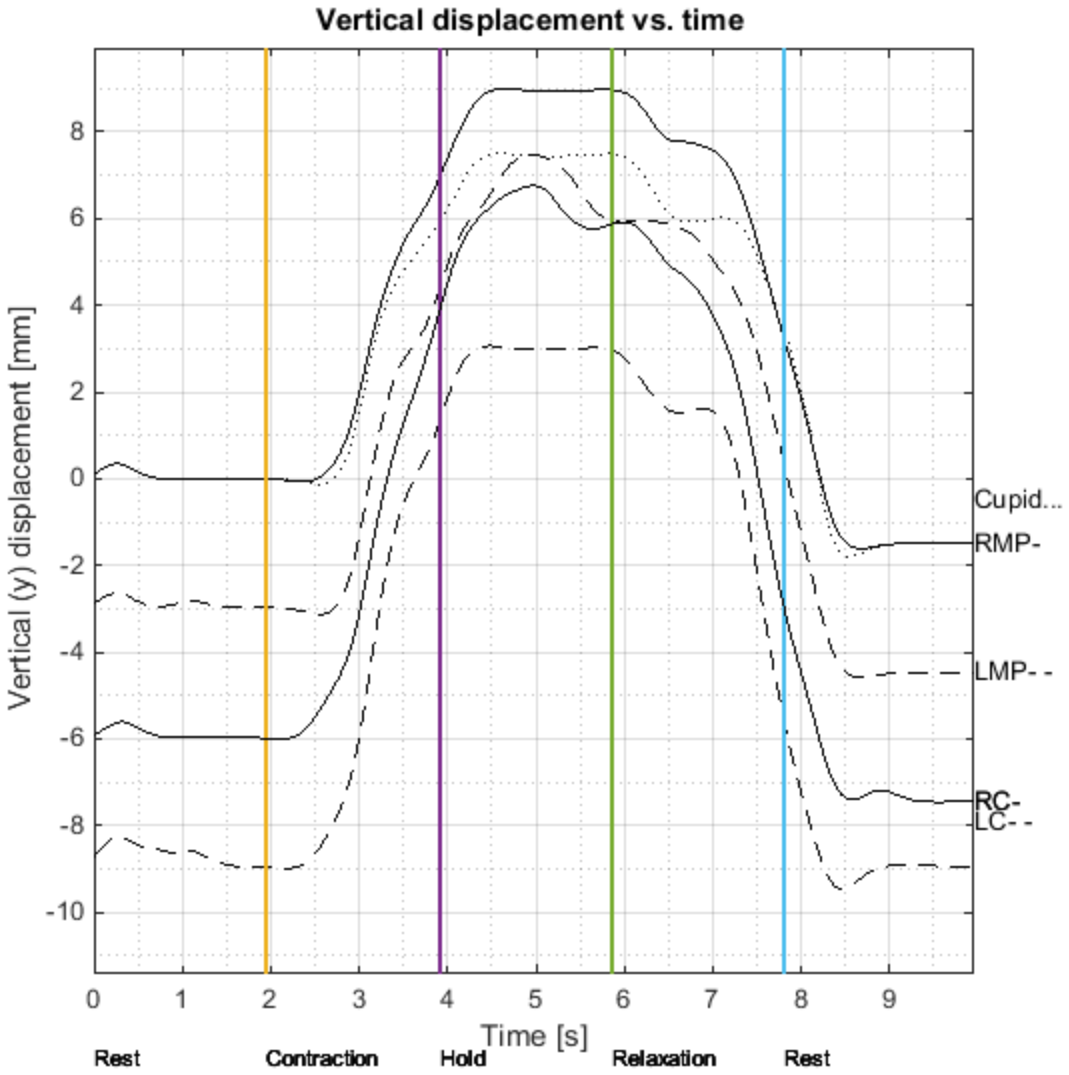
Maximum Extensions:

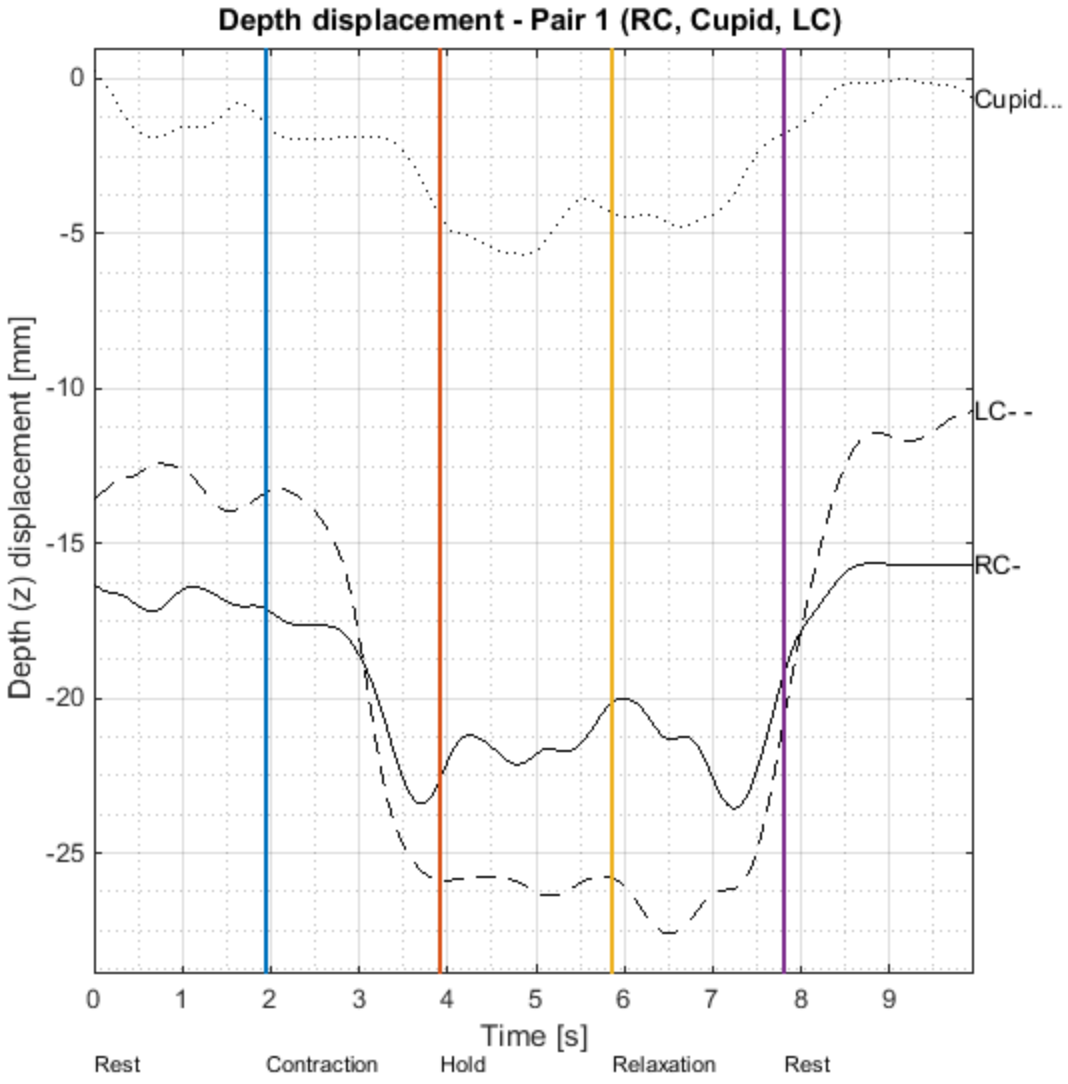
	Max. Extensions [mm]
Commissure R	18.0
Midpoint R	14.1
Cupid	8.5
Midpoint L	17.4
Commissure L	24.2

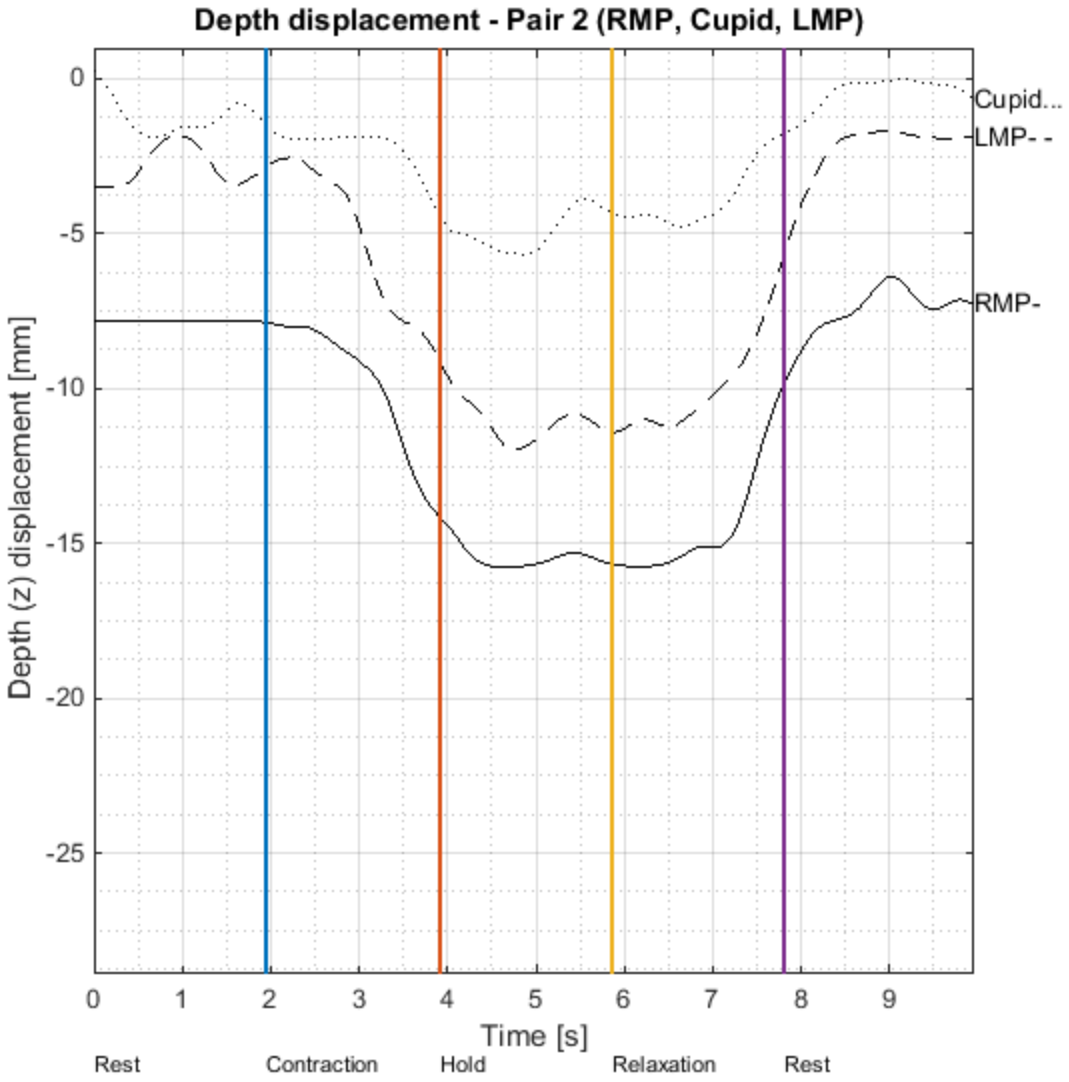
Symmetry index:
Commissures: 0.80
Midpoints: 0.79

Dynamic analysis

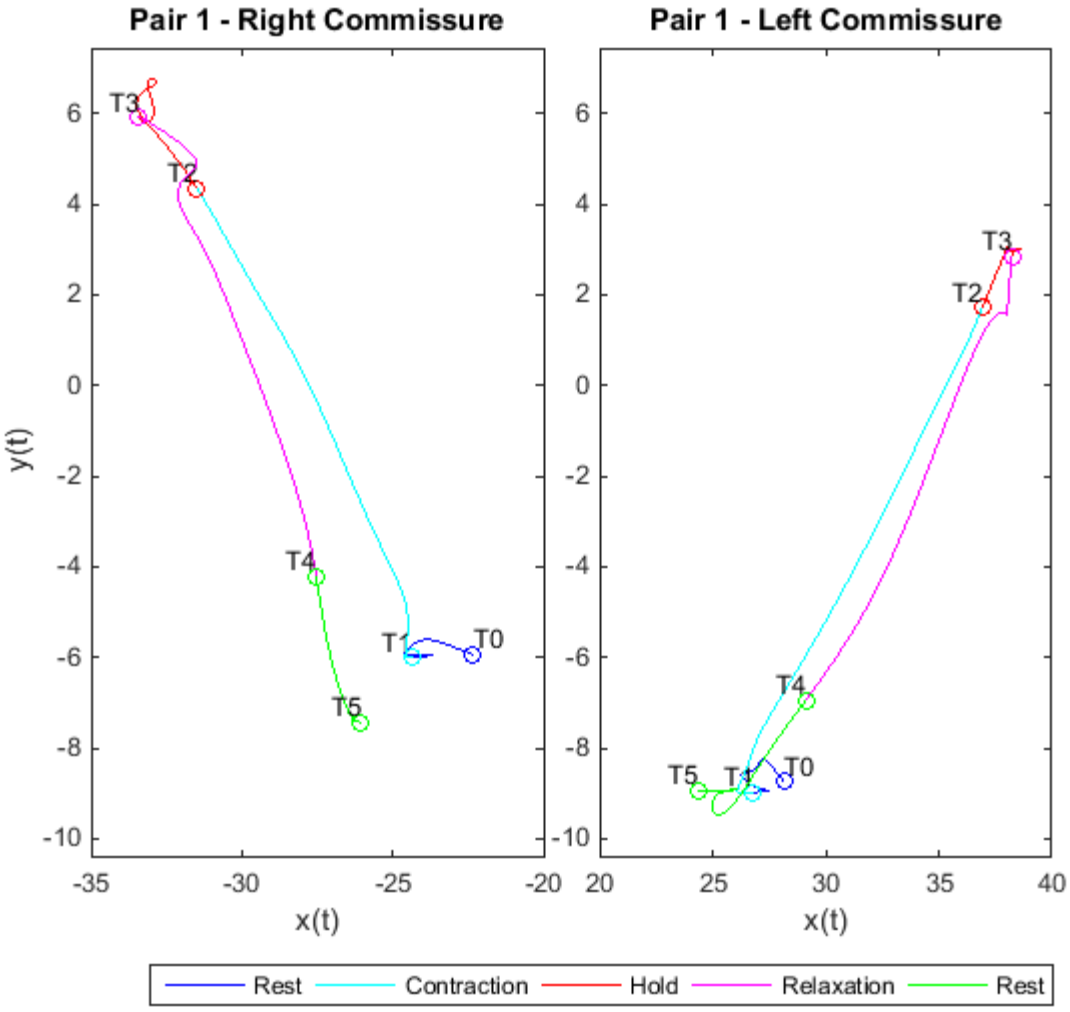


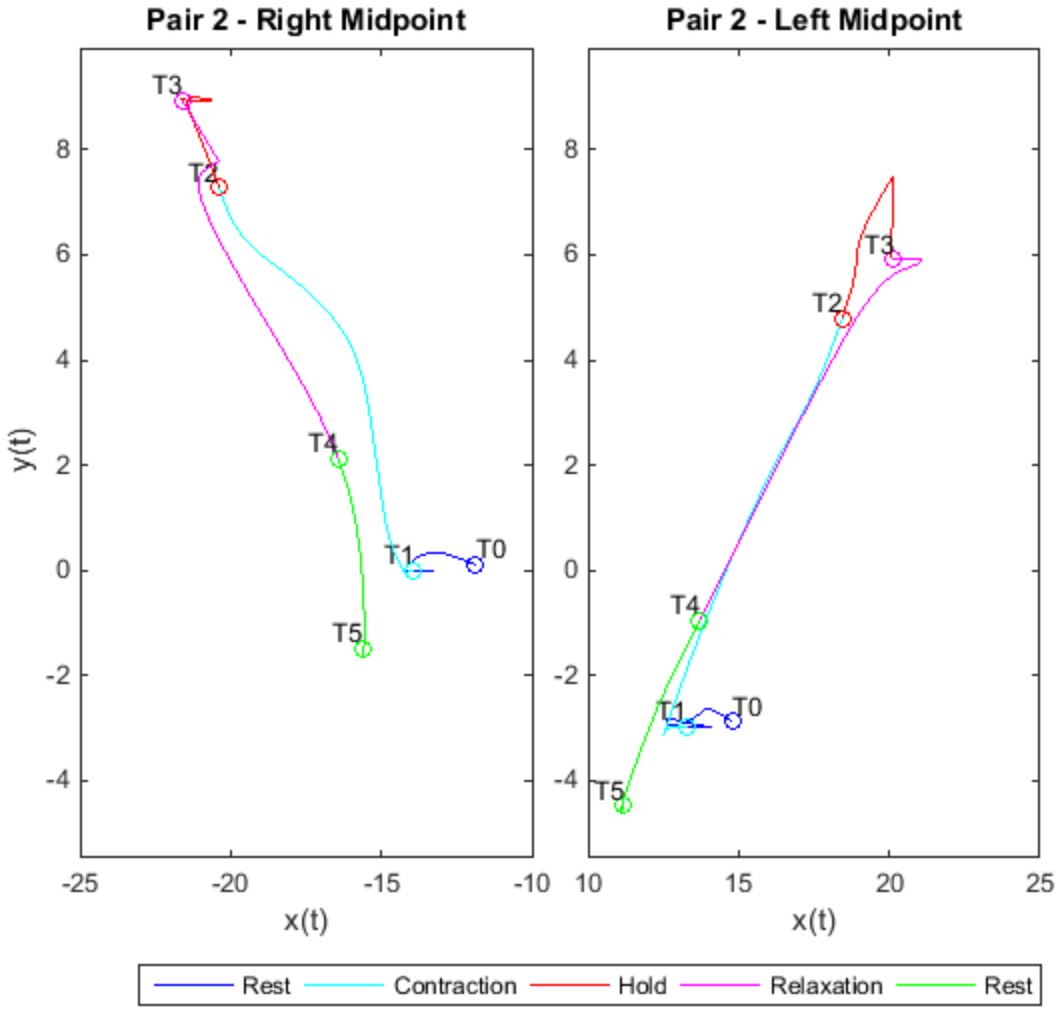

Vertical displacement vs. time
Vertical (y) displacement [mm]
8
6
4
2
0
-2
-4
-6
-8
-10
0
1
2
3
4
5
6
7
8
9
Time [s]
Rest
Contraction
Hold
Relaxation
Rest
Cupid...
RMP-
LMP- -
RC-
LC- -

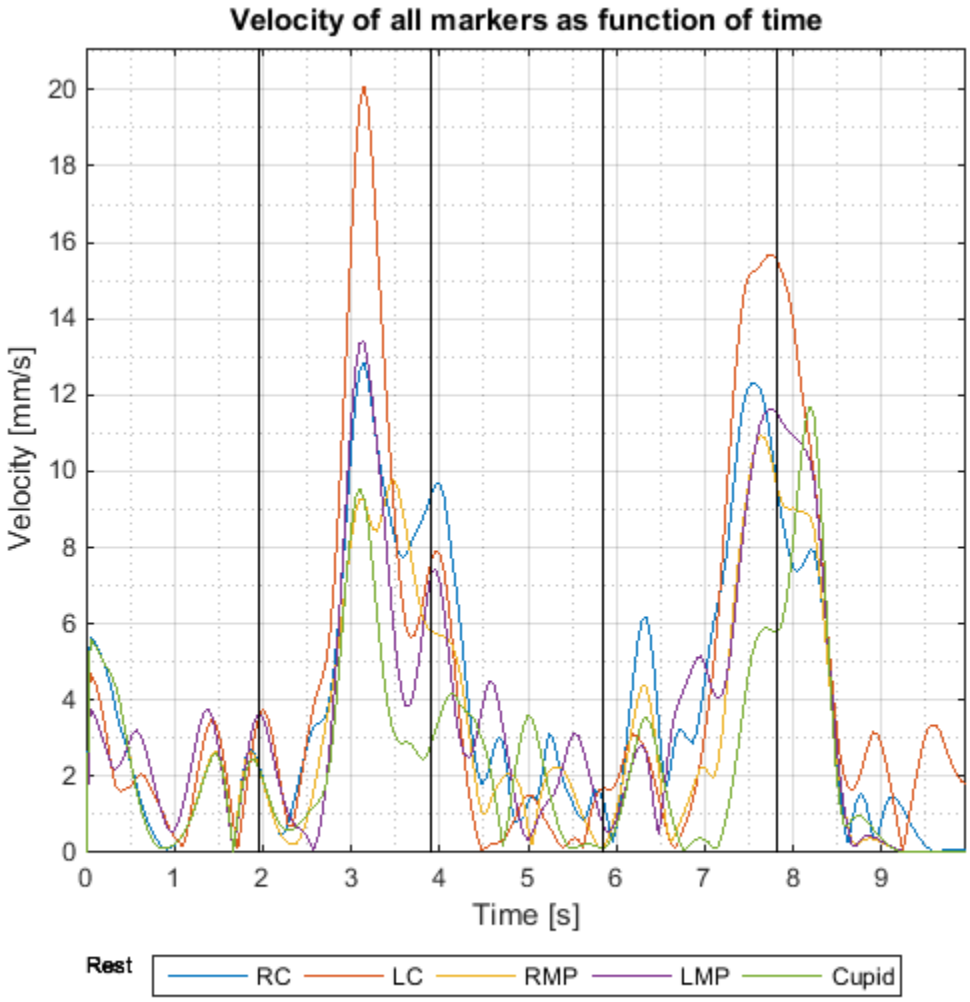

Depth displacement - Pair 1 (RC, Cupid, LC)
Depth (z) displacement [mm]
0
-5
-10
-15
-20
-25
0
1
2
3
4
5
6
7
8
9
Time [s]
Rest
Contraction
Hold
Relaxation
Rest
Cupid...
LC- -
RC-



Anatomical pairs' pathways:


Pair 1 - Right Commissure
Pair 1 - Left Commissure
y(t)
x(t)
x(t)
T0
T1
T2
T3
T4
T5
Rest
Contraction
Hold
Relaxation
Rest


Pair 2 - Right Midpoint
Pair 2 - Left Midpoint
T0
T1
T2
T3
T4
T5
y(t)
x(t)
x(t)
Rest
Contraction
Hold
Relaxation
Rest



Synchronization index:

	x Coord.	**y Coord.**
Pair 1 - RC& LC	-0.55	0.85
Pair 2 - RMP& LMP	-0.43	0.91

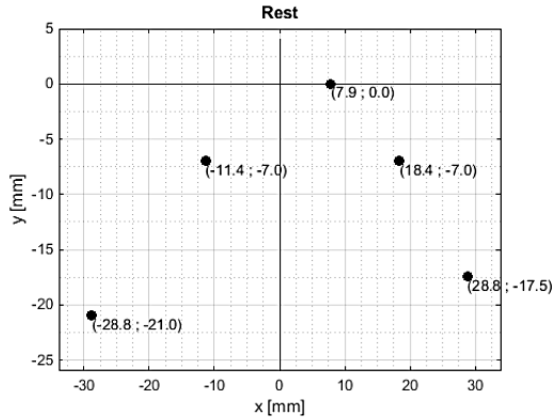
Facegram

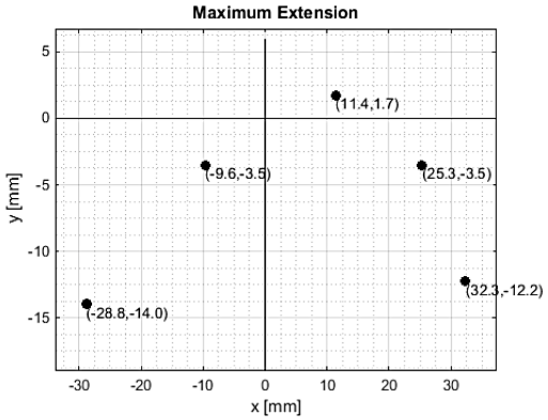
FACEGRAM.

Name: patient1

Sex: Female Age: -- Date: 27-Jun-2015

Static analysis





Maximum Extensions:

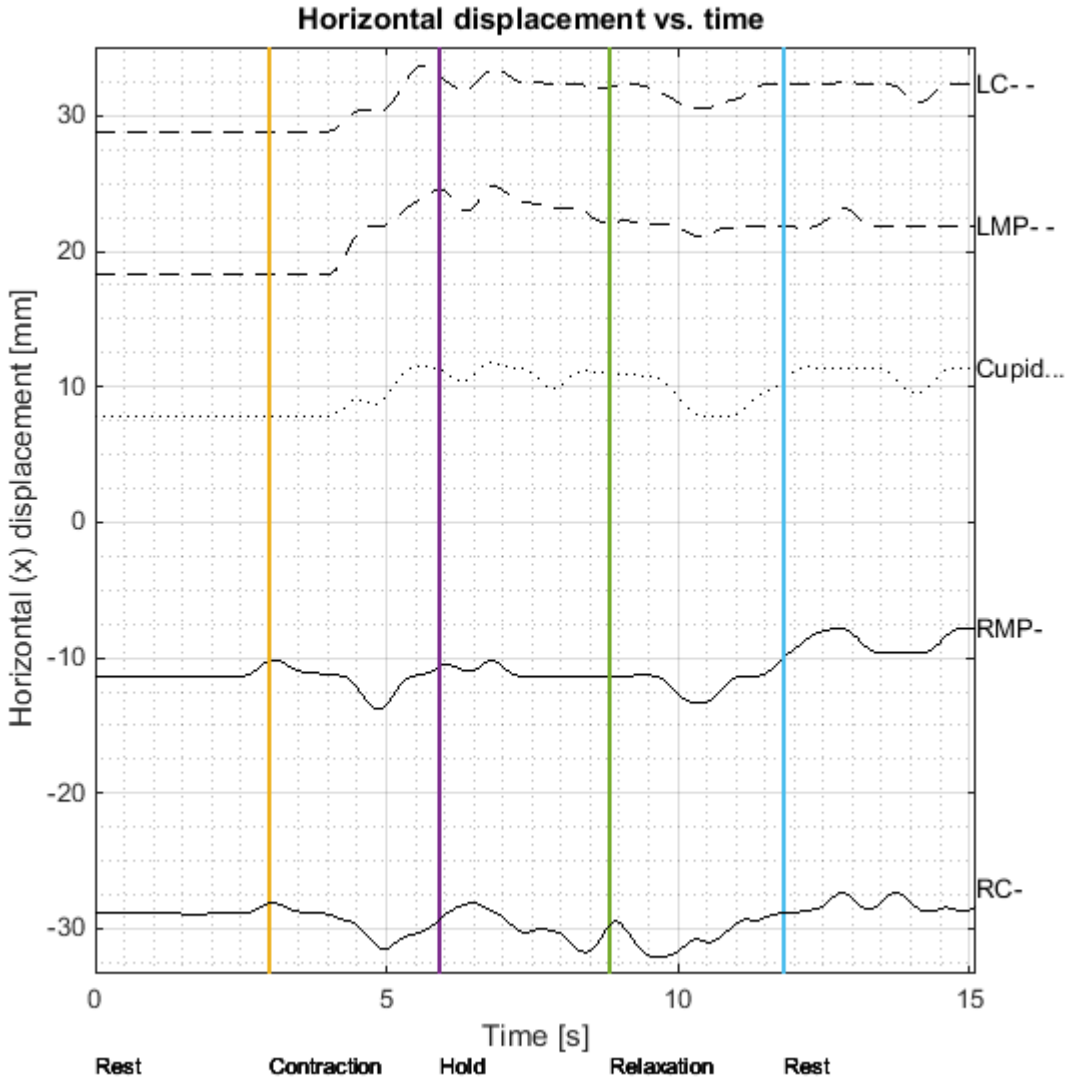
	Max. Extensions [mm]
Commissure R	8.5
Midpoint R	6.8
Cupid	7.0
Midpoint L	12.1
Commissure L	10.3

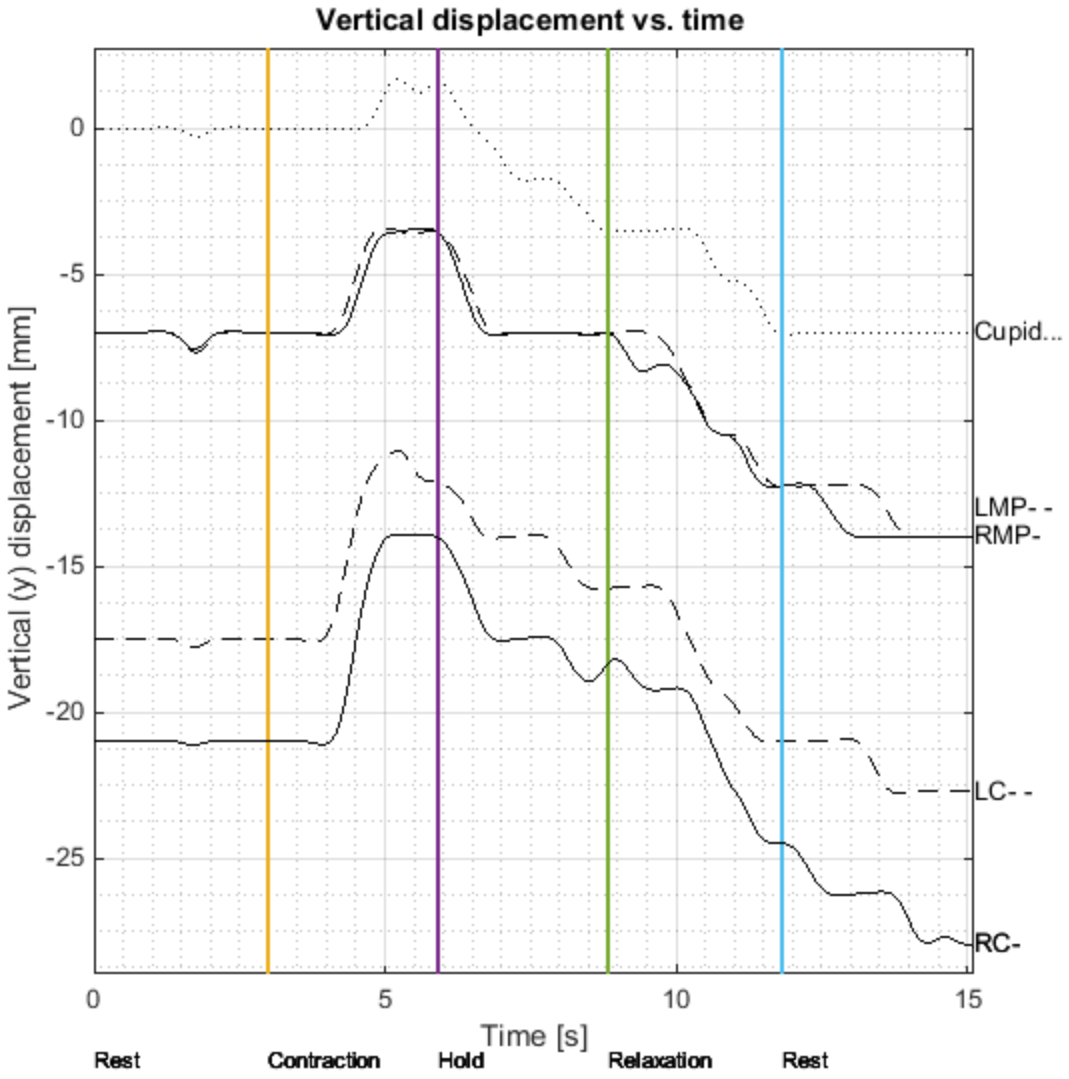
Symmetry index:

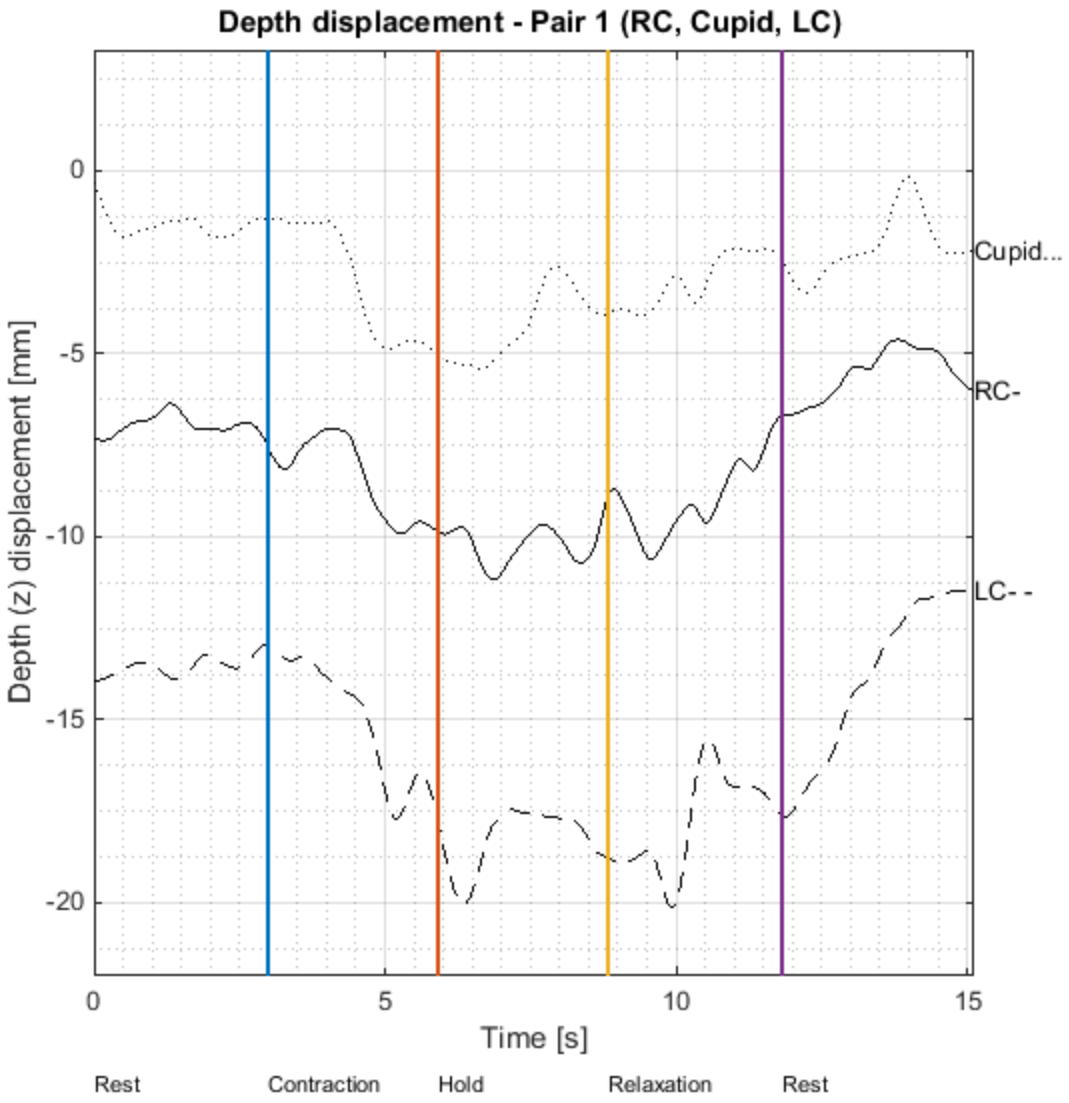
Commissures: 0.78

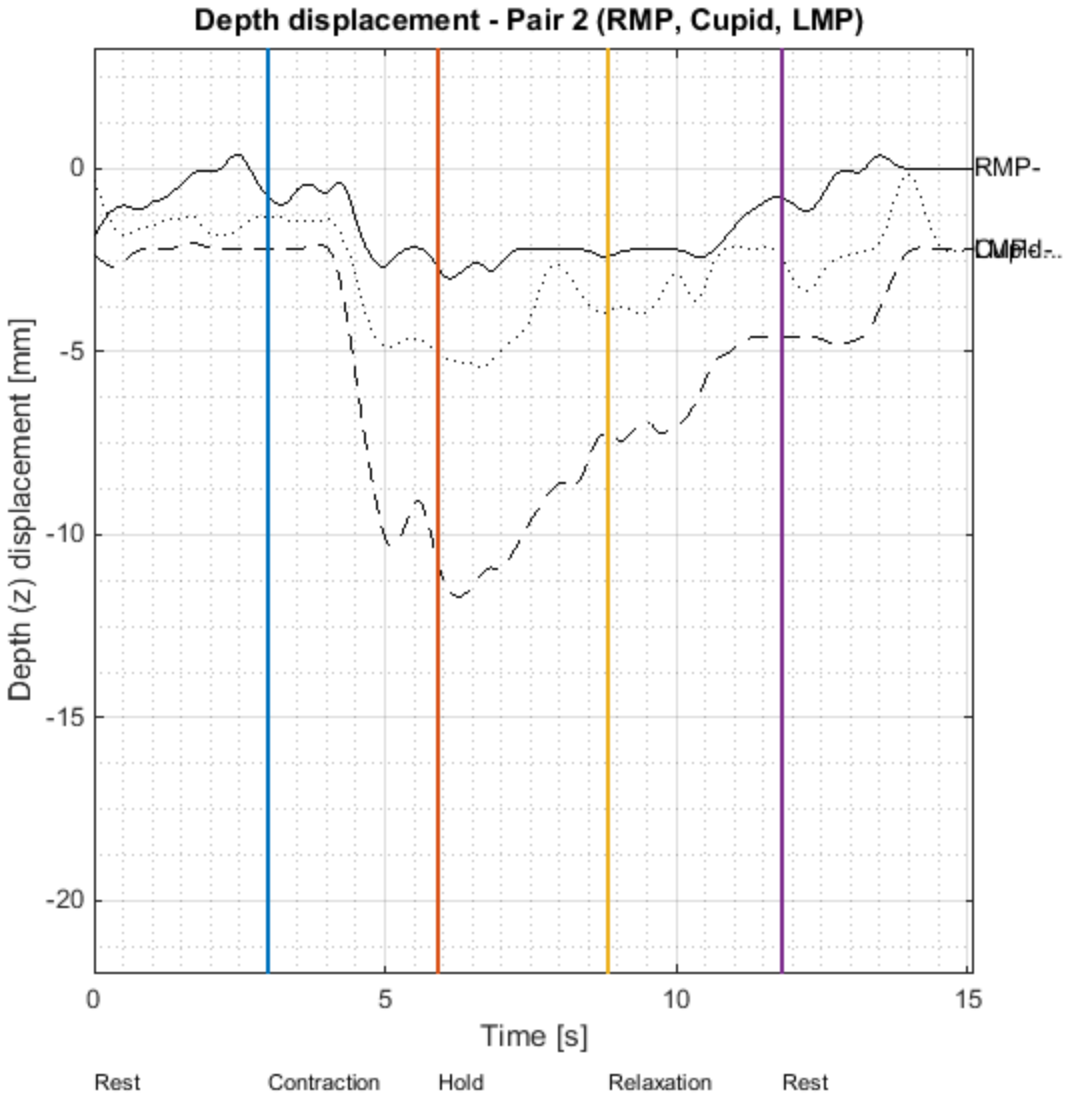
Midpoints: 0.34

Dynamic analysis

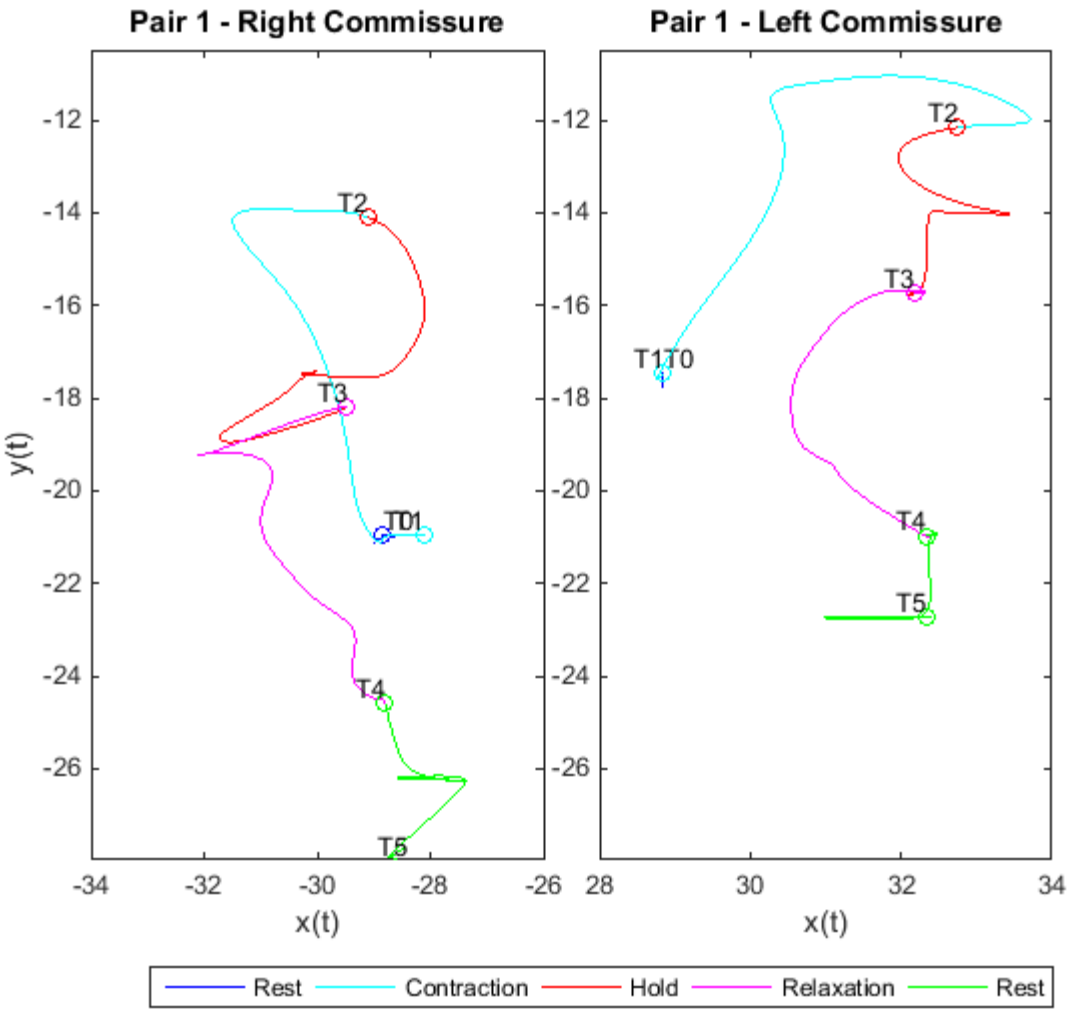


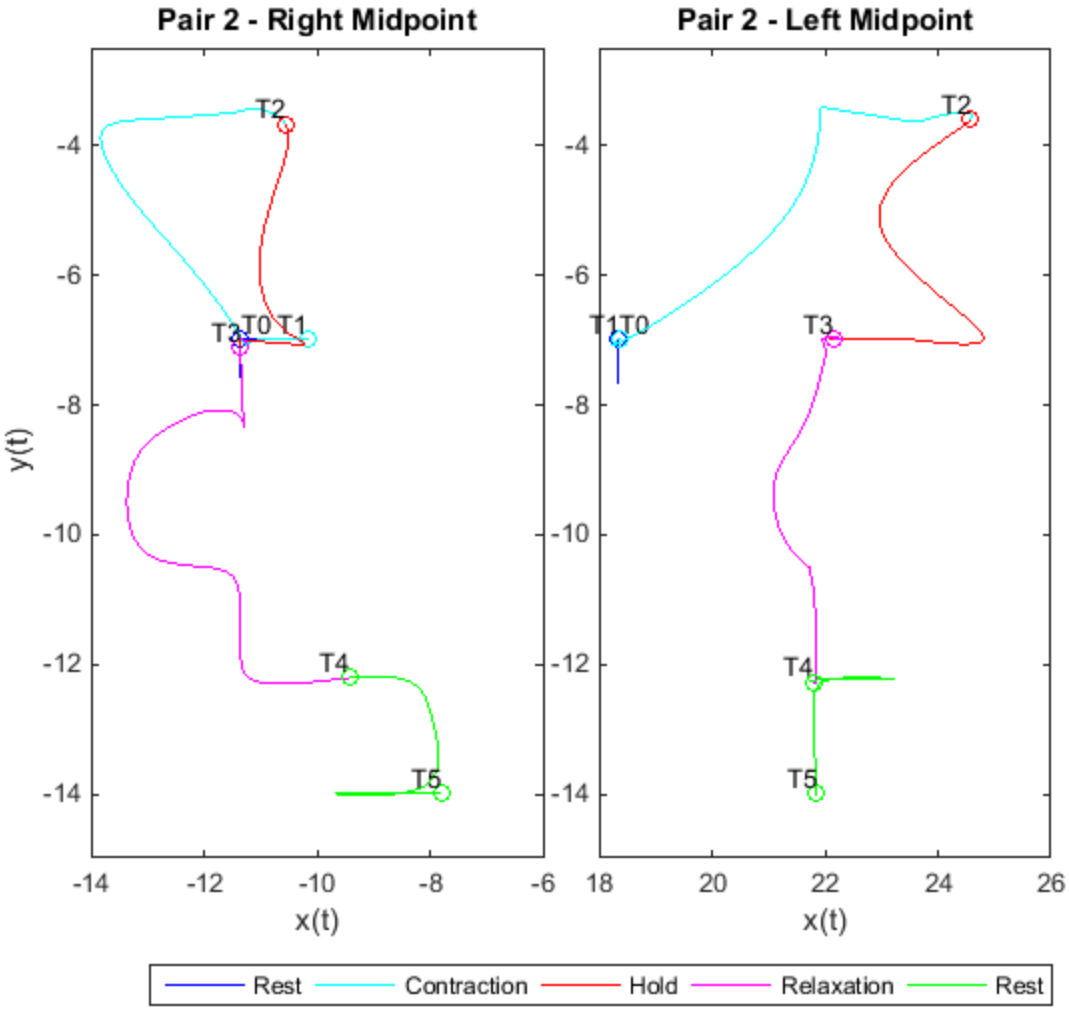

Vertical displacement vs. time
Vertical (y) displacement [mm]
0
-5
-10
-15
-20
-25
0
5
10
15
Time [s]
Cupid...
LMP- -
RMP-
LC- -
RC-
Rest
Contraction
Hold
Relaxation
Rest

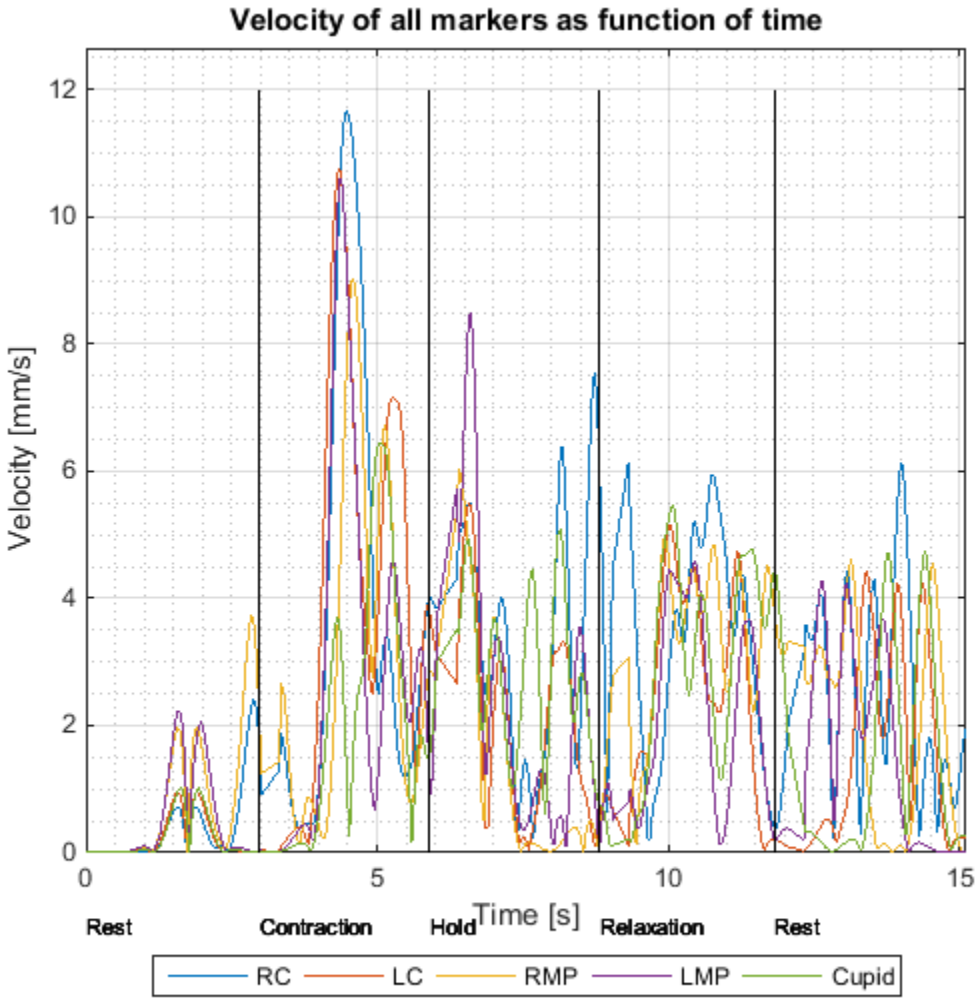

Depth displacement - Pair 1 (RC, Cupid, LC)
Depth (z) displacement [mm]
0
-5
-10
-15
-20
0
5
10
15
Time [s]
Cupid...
RC-
LC- -
Rest
Contraction
Hold
Relaxation
Rest



Anatomical pairs' pathways:


Pair 1 - Right Commissure
Pair 1 - Left Commissure
y(t)
x(t)
x(t)
T0
T1
T2
T3
T4
T5
Rest
Contraction
Hold
Relaxation
Rest


Pair 2 - Right Midpoint
Pair 2 - Left Midpoint
T0
T1
T2
T3
T4
T5
y(t)
x(t)
-4
-6
-8
-10
-12
-14
-14
-12
-10
-8
-6
18
20
22
24
26
Rest
Contraction
Hold
Relaxation
Rest



Synchronization index:

	x Coord.	**y Coord.**
Pair 1 - RC& LC	-0.01	0.88
Pair 2 - RMP& LMP	0.13	0.82

Facegram

Appendix E

Medical Statement

On the next page is illustrated the statement written by the specialists from *Plastic Surgery, Reconstructive and Maxillofacial* center, at *Hospital São João*, Porto. This document emphasizes the importance that *Facegram* is likely to have in the clinical environment.

SÃO JOÃO
HOSPITAL

CIRURGIA PLÁSTICA, RECONSTRUTIVA, ESTÉTICA E MAXILO-FACIAL (U.A G de Cirurgia)
HOSPITAL DE SÃO JOÃO, E.P.E.
Director- Álvaro Silva

O trabalho desenvolvido e incluído na tese de mestrado da estudante Ana FilipaGerós, e superiormente orientado pelo Professor Paulo de Castro Aguiar, assenta numa parceria entre a Faculdade de Ciências e a Faculdade de Medicina da Universidade do Porto.

O *Facegram* apresenta-se como uma ferramenta objetiva, precisa e prática para a avaliação clínica de doentes com desfiguramento facial, nomeadamente no contexto de paralisias faciais, queimaduras, lesões pós-traumáticas, doenças neurológicas, permitindo a avaliação espaço-temporal de pontos de referência na face. Permitirá reconciliar e comparar os doentes no pré e pós-operatório, avaliar a eficácia de terapêuticas médicas distintas (reabilitação, reanimação facial com técnicas estáticas ou dinâmicas), acompanhar a recuperação dos doentes com acidente vascular cerebral (AVC), paralisia de Bell, entre outras aplicações.

O registo gráfico é de fácil interpretação e ao mesmo tempo muito informativo, pelo que o registo da dinâmica facial e a informação relativa a assimetrias no domínio do espaço e tempo será muito relevante na prática clínica.

29/06/2015

Ricardo Horta

Faculdade de Medicina da Universidade do Porto

Serviço de Cirurgia Plástica, Reconstrutiva e Maxilo-Facial- Centro Hospitalar São João, Porto.