UNIVERSITY OF TWENTE.

Radboudumc

Technical medicine M.Sc. Thesis

3D imaging for the prediction of a difficult airway

Author:
T.G.J. Loonen

Technical Supervisor UT:
Dr. Ir. F. vd Heijden

Medical Supervisor UMC:
Prof. G.J. Scheffer

Technical Supervisor UMC:
Dr. T.J.J. Maal

Project Supervisor:
Dr. M. Groenier

7th June 2017

Foreword

In the thesis at hand I, Tom Loonen, present my conducted research at the department of Anaesthesiology and the 3D lab of the Oral and Maxillofacial surgery department at the Radboud UMC Nijmegen. This thesis is made as a completion of the master's Medical Imaging and Interventions in Technical Medicine from the University of Twente located in Enschede. Besides the work for this thesis I also participated in other smaller projects and activities this year. A short summary of these projects and activities is in given in appendix A. I hope my work will continue and will have a significant impact on healthcare.

Several persons have contributed academically, clinically, practically and with support to this master thesis. I would therefore firstly like to thank the head of the department of anaesthesiology and my medical supervisor professor Scheffer for giving me the opportunity to optimize the prediction of a difficult airway with 3D imaging. In particular, the freedom I have had to improvise on the various possibilities for achieving this goal.

Secondly I would like to thank dr. Thomas Maal, head of the 3D Lab Nijmegen, for providing me with the opportunity to join the 3D Lab during my thesis. Thomas has given me much confidence to work independently. In addition, he also involved me in many other projects, which made me feel a valued colleague.

Also, I would like to thank Jene Meulstee, with which I worked the most for the development of the automatic landmarking algorithm in chapter 3.

I would also like to thank dr. Loes Bruijstens for her expertise and knowledge in the field of anaesthesiology and difficult airway issues. She taught me to think like a clinician and to perform airway related clinical activities like, mask ventilation and intubation.

I also thank all other members of the 3D lab Nijmegen for the interesting discussions, useful recommendations and for all the fun we had the last year. And of course, for introducing me in the world of cryptocurrency.

Furthermore, I would like to thank dr. ir. van der Heijden at University of Twente for his help throughout the entire technical and writing process. I would also like to thank dr. Groenier at University of Twente for support throughout the last two years of my master period.

Special and final thanks go out to my parents, girlfriend, my brother and his wife who have always been there for me. They made it possible for me to complete an extra 6 year of studying after an already 4 years of mechanical engineering study. Without them it would have been impossible to complete a total of 10 years studying with all diplomas and outstanding results. Thanks.

Study outline and Summary

All patients undergoing general anaesthesia are evaluated pre-operative for the prediction of a difficult intubation. These intubations can become complicated due to a difficult airway and can result in a "can't intubate can't oxygenate" (CICO) situation with serious consequences, like brain damage and death. The prediction of possible problems allows the anaesthesiologist to change plans or summon help and/or alternative equipment. Poor airway assessment and the failure to change strategy according to the assessment contributes to more airway related complications. The incidence of difficult intubation (1,9%) and complications is relatively small, but with high number of general anaesthetics given (yearly 10,000 at the RadboudUMC), and a possible catastrophic outcome, it is still an important field of research. Difficult airway prediction and airway management is studied for many years and a lot of methods are developed to predict a difficult airway. Unfortunately, all prediction parameters and methods are complicated or little sensitive and little specific [1]. An easy and accurate prediction of a difficult airway can prepare anaesthesiologists to prevent serious complications and stressful situations during general anaesthesia. Therefore, in this study we used 3D imaging techniques to improve the prediction of difficult intubation with two methods. One method is the development of a machine learning prediction model and is further subdivided in two studies: (i) literature study about documentation parameters (Chapter 2) and (ii) Automatic landmarking development for 3D stereophotographs (Chapter 3). The second method and third study (iii) is the use of virtual laryngoscopy (Chapter 4).

Baseline study (Chapter 2)

For the development of a computerised prediction model, data has to be divided in two groups, easy-to-intubate and difficult-to-intubate. Therefore, intra-operative documentation of these parameters is necessary. In chapter 2 a literature study is started about difficult intubation definitions and the parameters necessary for documentation. Pre-operative bed-side test parameters for prediction of difficult intubation are investigated also, to check the nowadays used method. This resulted in the addition of one parameter pre-operative. The anaesthesiologist needs to document if the patient is predicted difficult or not. With this parameter, it is possible to compare a new method with the prediction performance of the anaesthesiologist. Intra-operative the following parameter were added besides the Cormack & Lehane grade: number of intubation attempts, number of experienced operators, change of primary plan and the use of extra materials or techniques. A single parameter or combinations of parameters can be used to define if an intubation was easy or difficult. With these changes in documentation, groups can be made for the development of a computerised prediction model and changes of performance in airway management can be monitored.

DTM landmarking (Chapter 3)

The prediction model will be trained with features derived from 3D stereophotographs and the classification based on parameters from chapter 2. The input features of the model will be landmarks on 3D stereophotographs and therefore a study was started for the automatic annotation of these landmarks on the images. This study uses a facial template with the landmarks manually annotated. The template is matched on a patient's 3D stereophotograph with a non-rigid coherent point drift algorithm and is called a deformable template matching (DTM) method. The positions of the automatically detected landmarks are verified by comparison with manually placed landmarks by three observers (ground truth). The normalized distance-error (NDE) and accuracy of the presented method are used for performance measures of the method. The method is also compared to two commercially available automatic landmarking software packages. The method is verified on 10 male and 10 female volunteers. The overall NDE and accuracy of the DTM method were significantly better compared to the commercially available systems (p <0.001). Improvements of the method can be made at the eye landmarks by using a texture based instead of a shape based algorithm. This study showed that the presented DTM algorithm is capable of extracting features out of 3D stereophotographs for the training of a difficult intubation prediction model with machine learning.

Virtual laryngoscopy pilot study (Chapter 4)

Another method for the improvement of the prediction of a difficult intubation is to make use of medical imaging already available for specific patients. This study investigated the added value of a virtual laryngoscopy with CT image data on patients with a suspicion of a difficult airway. Patients undergoing head and neck surgery were frequently reported (39%) in complication cases in the UK [2]. Due to pathology or previous surgery in this region a large group of these patients have some kind of medical imaging of this area. In this study the already available image data is used to make a virtual laryngoscopy. The data is evaluated with anaesthesiologist to investigate the usefulness of this new technique and the possible extra features necessary for implementation. A total of four virtual laryngoscopies are made and are discussed with three different anaesthesiologists. However, some extra features need to be added to the algorithm to improve functionality, the results were positive. A big advantage of the virtual laryngoscopy is the 3D view and thereby a good understanding of the anatomy. The study is continued by development of a specific software program with all extra functionalities. After completion of the software a larger study can be started to draw solid conclusions about the added value of virtual laryngoscopy for the prediction of a difficult airway.

Conclusions

Based on these two methods and the three conducted studies it can be concluded that the search for better difficult airway prediction methods is still an ongoing process. The presented two methods are new for the department of anaesthesiology and both have promising results for the future. The methods need further investigation but machine learning models based on 3D stereophotographs and virtual laryngoscopy could have a significant impact on healthcare for the department of anaesthesiology.

Contents

List of Figures

List of Tables

List of Abbreviations

2D .... Two-dimensional
3D .... Three-dimensional
AUC .... Area-under-the-curve
BMI .... Body mass index
CBCT .... Cone beam computed tomography
CICO .... ”can’t intubate can’t oxygenate”
CNN .... Convolutional neural networks
CPD .... Coherent point drift
CT .... Computed tomography
DL .... Direct laryngoscopy
DTM .... Deformable template matching
GPU .... Graphics processing unit
HU .... Hounsfield units
ICC .... Intraclass correlation coefficient
ICU .... Intensive care unit
IDS .... Intubation difficulty scale
IOD .... Inter-ocular distance
MRI .... Magnetic resonance imaging
NC .... Neck circumference
NDE .... Normalised distance-error
ROC .... Receiver operating characteristic
SAD .... Supraglottic airway device
SGD .... Supraglottic airway device
TMD .... Thyromental distance
TMJ .... Temporomandibular joint
ULBT .... Upper lip bite test
UT .... Ultrasound

Chapter 1

General introduction

All patients undergoing general anaesthesia are evaluated pre-operative for the prediction of a difficult intubation. The prediction is used to choose the optimal intubation technique and to prepare the anaesthesiologist for possible complications during ventilation of the patient. The so-called airway management is based on physical evaluation of anatomical features and the experience of the anaesthesiologist. Based on a British national audit, tracheal intubation was used to secure the airway in 38% of all general anaesthetic cases [2]. These intubations can become complicated due to a difficult airway. The incidence of a difficult airway depends on the definition and is explained below. A difficult airway can result in a "can't intubate can't oxygenate" (CICO) situation. Failed intubation and CICO are common causes during anaesthesia and account for 23% and 14%, respectively, of primary airway problems. The review panel of the national audit assessed the airway management as poor in 34% of all difficult airway cases under general anaesthetics [2].

The definition of a CICO situation according to Nagaro et al., is a situation in which both ventilation with a face mask and bag, despite the use of oral or nasal artificial airways, and tracheal intubation with direct laryngoscopy is difficult, and life-threatening events cannot be avoided without additional airway interventions [3]. In this study the incidence of CICO in 60 university hospitals in Japan, in one year, is 0.017% ($26/151,900$). According to Eindhoven et al. difficult intubation and difficult mask ventilation during surgery occurs in 0.37% of all patients [4]. According to Cook et al. the incidence of CICO and airway related complications are, 1 per 167,000 and 1 per 22,000, respectively [2]. However the incidence is rare, complications related to airway management are catastrophic, such as: death, brain damage, emergency surgical airway and intensive care unit (ICU) admission. A low likelihood with a very high impact, results in a high-risk factor (table 1.1 encircled position). Therefore, improvement of airway management is an important component in improving patient safety. The start for better airway management is the improvement of prediction of a difficult airway. The prediction of problems with airway management allows the anaesthesiologist to:

- Summon help
- Assemble alternative equipment
- Undertake airway management in a more equipped area
- Secure the airway by awake intubation
- Undertake cricothyroidotomy or tracheostomy
- Offer regional or local anaesthesia instead of general anaesthesia
- Postpone elective surgery

Poor airway assessment and failure to change strategy after a potential difficult airway is detected, contributes to poor airway outcomes [2]. Difficulties with predication of a difficult airway or securing the airway are compounded by inter-observer variability, inaccuracies in measured airway variables at the bedside, and the wide range of airway devices, each having a different impact on the airway [5].

Table 1.1: Risk factors. The encircled position shows the Risk factor for a difficult intubation

Risk factors		Impact			
		Low	Moderate	High	Very High
Likelihood	Likely	Moderate	High	High	High
	Moderate	Moderate	Moderate	High	High
	Unlikely	Low	Moderate	Moderate	High

The incidence of difficult intubation depends largely on the definition. The definitions for difficult intubation and difficult airway are inconsistent and vary between anaesthesiologists and hospitals [6]. Three different definitions are (i) poor laryngoscopic view, grade 3 and 4 by the Cormack and Lehane scale [7], (ii) three or more attempts, and (iii) failure of intubation with direct laryngoscopy. According to these definitions Rose et al. found the following incidence rates: (i) 10.1%, (ii) 1.9% and (iii) 0.1% [6]. Also, combinations of these three definitions and other definitions are used.

The incidence of difficult intubation and complications is relatively rare, but with high number of general anaesthetics given, and a possible catastrophic outcome, it is still an important field of research. Number of general anaesthetics which require tracheal intubation in the Radboud UMC is about 10,000 each year. Difficult airway prediction and airway management is studied for many years and a lot of methods are developed to predict a difficult airway. Based on these studies also many recommendations and guidelines are developed. Unfortunately, all prediction methods are complicated or little sensitive and little specific [8]. An easy and accurate prediction of a difficult airway can prepare anaesthesiologists to prevent serious complications and stressful situations during general anaesthesia.

To improve airway management we want to improve the difficult airway prediction and therefore two methods for further investigation were defined. One method is the development of a prediction model and is further subdivided in two studies: (i) literature study about documentation parameters and (ii) Automatic landmarking development for 3D stereophotographs. The second method and third study (iii) is the use of virtual laryngoscopy. This results in a total of three studies. The hierarchy is shown in figure 1.1 and the studies are further explained below.

(i) Documentation parameters for baseline study (Chapter 2)

Pre-operative difficult airway prediction parameters and some information about the difficulty of intubation are documented. However, this information is not used to evaluate and monitor the prediction performance and difficult intubation incidence at the Radboud UMC. No information is available about accuracy of difficult airway prediction, incidence of difficult intubation and incidence of a difficult airway. The prediction is made during a pre-operative consult by an anaesthesiologist. Therefore, a baseline study is needed to compare performance of the anaesthesiology department with the performance known in literature. For this study, some adjustments have to be made in the documentation method of the airway prediction and

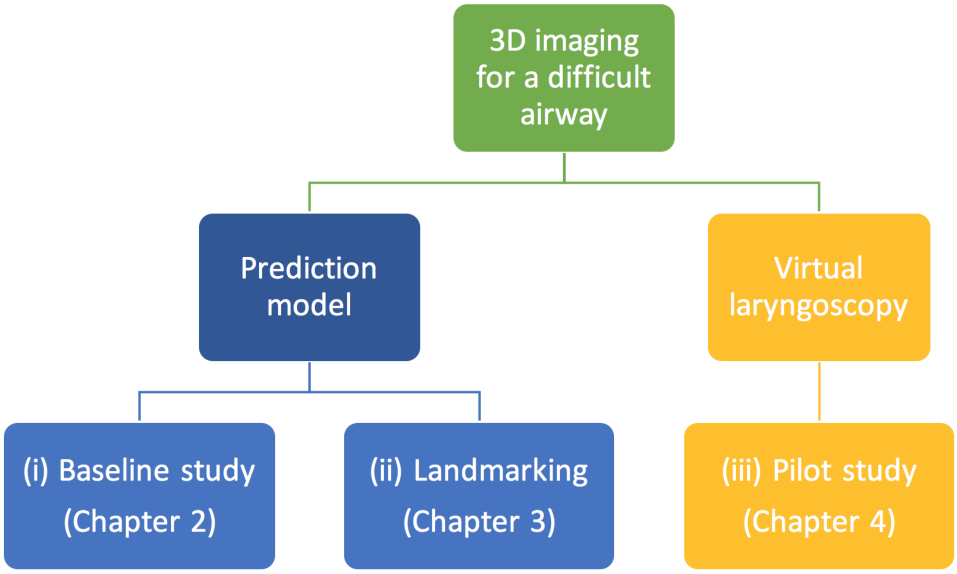


Figure 1.1: Study hierarchy and relation between the sub studies and corresponding chapters.

the preoperatively monitored parameters concerning intubation and airway. With the extra data, it is also possible to monitor performance after the implementation of changes in airway management. During this study, literature research will be done to make a decision about the extra parameters that should be stored in the patient's medical record. These changes will also be put into practice during this study.

The research question for this path was: Which changes need to be made in the documentation of airway management at the Radboud UMC, in order to gather data for a baseline study and monitoring of the performance of airway management?

(ii) Automatic landmarking on 3D facial stereophotographs based on deformable template matching (Chapter 3)

Recent studies showed that analysis of parameters derived from facial photographs are useful for the prediction of difficult intubation [1,9]. These studies used two-dimensional (2D) photographs of patients. One study used only the natural head pose and the other study used also positions to measure mouth opening and neck extension. Features derived from these photographs were used to make a prediction model with machine learning. Based on these studies the Radboud UMC started a pilot study to make a computer model based on three-dimensional (3D) facial stereophotographs with different head positions. This study was performed by the author during a technical medicine master internship of ten weeks. The hypothesis was that 3D stereophotographs instead of 2D could improve accuracy because there is a possibility to also measure distances instead of only ratio's and angles. This pilot study contained too little subjects to draw solid conclusions about the accuracy of the method but it showed the feasibility and potentials of the method. The purpose of the study performed in this path was to continue the development of a computerised prediction model. For the training of a computerised model, data acquisition for more 3D stereophotographs is needed. The analysis of these stereophotographs takes a lot of time and the first step is to automate this analysis. In this study, an automatic analysis method was developed for acquiring of landmarks on the 3D stereophotographs in different positions. The used method, is a deformable template matching (DTM) method. Later, the automatic landmarks can be used to calculate features which are used as input for machine learning. The new developed method was verified with the golden standard for landmark annotation.

The research question for this path was: What is the performance of a new developed deformable template matching landmark detection algorithm for use with 3D stereophotographs?

(iii) Added value of virtual laryngoscopy in patients suspected for a difficult airway (Chapter 4)

Patients planned for general anaesthesia are evaluated pre-operative for the prediction of a difficult intubation. Some difficult intubation predictors are: previous surgery in the head and neck region, space occupying lesions, large deformations in the region and radiotherapy. The prediction is mostly made based on assessment of the head and neck region from outside or history of treatments. This information lacks knowledge about the airway and surrounding tissue itself and therefore an accurate prediction of the difficulty of intubation is not possible. Due to pathology or previous surgery in this region a large group of these patients have some kind of medical imaging of this area. CT, MRI or CBCT data could be used to make a 3D virtual laryngoscopy of the airway and therefore has the potential to be of added value for the prediction of a difficult airway. Ahmad et al. described the potential benefit of virtual laryngoscopy but did not support this statement with any kind of research [10]. Cuendet et al. did some research about the correlation between virtual and conventional endoscopy. They described a good correlation and a benefit for virtual endoscopy because it is non-invasive, examination of the images is relatively easy, it is cheaper and it is useful for training [11]. Besides these benefits of virtual endoscopy, the 3D image data can also be used to make measurements and gather quantitative data about airway obstructions. In this study, the potential added value of virtual endoscopy and measurements based on 3D image data, for the prediction of difficult intubation and help with planning of airway management for suspected patients was investigated.

The research question for this path was: Is a virtual laryngoscopy useful for airway planning in patients who are suspected for a difficult airway, due to anatomical changes in or near the airway?

Chapter 2

Baseline study

Pre-operative difficult airway prediction parameters and some information about the difficulty of intubation are documented. However, this information is not used to evaluate and monitor the prediction performance and difficult intubation incidence at the Radboud UMC. No information is available about accuracy of difficult airway prediction, incidence of difficult intubation and incidence of a difficult airway. This information is mainly extrapolated from Danish, British or American research. No Dutch data is known about this subject. The prediction is made during a pre-operative consult by an anaesthesiologist. Therefore, a baseline study is needed to compare performance of the anaesthesiology department with the performance known in literature. For this study, some adjustments have to be made in the documentation method of the airway prediction and the pre-operatively monitored parameters concerning intubation and airway. With the extra data, it is also possible to monitor performance after the implementation of changes in airway management. During this study, literature research will be done to make a decision about the extra parameters that should be stored in the patient's medical record. These changes will also be put into practice during this study.

The research question for this path was: Which changes need to be made in the documentation of airway management at the Radboud UMC, in order to gather data for a baseline study and monitoring of the performance of airway management?

2.1 Methods and Materials

In order to answer the research question, the research was split in two, each with their own sub question. There are two main time periods of documentation concerning airway management. These periods are at the pre-operative consult and during surgery (intra-operative). At the pre-operative consult the documentation is about the prediction and planning regarding airway management. At surgery, this is about documenting the method and materials used to secure the airway. Therefore, the question was split into these two periods: (i) Which changes need to be made in the documentation of airway management at the pre-operative consult? (ii) Which changes need to be made in the documentation of airway management during surgery? To answer these questions a literature study was perform on the two different subjects.

To find parameters for documentation at the pre-operative consult, research was done about airway prediction parameters. We want to find the best airway prediction parameters for physical examination and compare them with the nowadays used parameters at the Radboud UMC. If there are big differences in the clinic with the literature, changes will be made in the workflow. The best and most practical prediction parameters will be used as support for the prediction of difficult intubation and will be documented also.

To find intra-operative parameters, research was done about the definition of a difficult intubation and a difficult airway. This is necessary for categorising of patients in the correct group, easy-to-intubate or difficult-to-intubate.

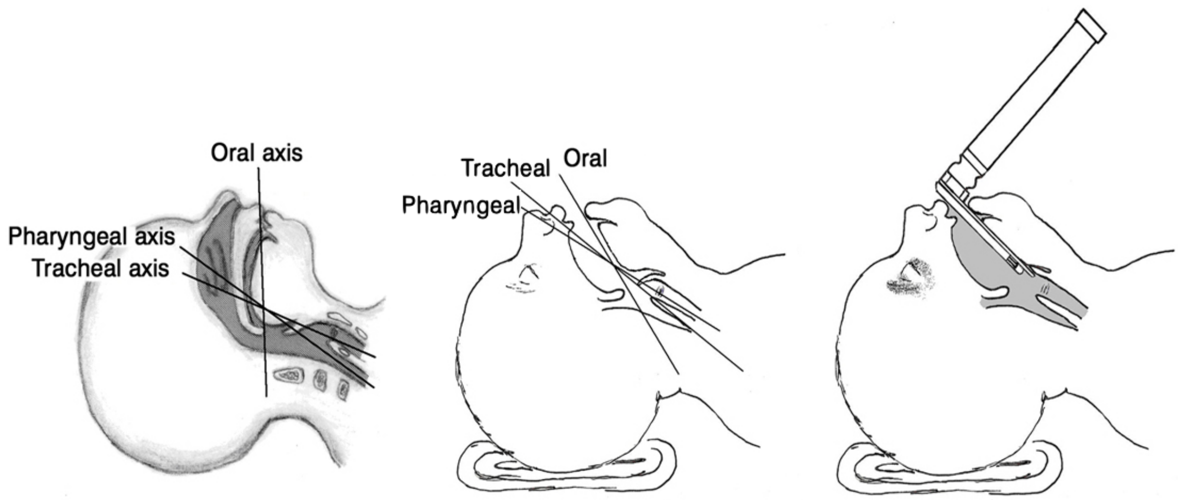
2.2 Results

2.2.1 Pre-operative parameters

In the section below the literature research about the predictive parameters for difficult intubation are summarised and the changes made to the documentation are discussed. Firstly, general information about direct laryngoscopy (DL) and the structures involved in direct laryngoscopy are described. Based on this information the mainly used predictive parameters are described.

Theoretical background for direct laryngoscopy

Difficulty of intubation has a strong correlation with difficult direct laryngoscopy which is graded according to the Cormack and Lehane scale [7]. Difficult laryngoscopy does not always cause difficult intubation but in the vast majority of patients, this relation is correct [12]. Therefore, in this section we emphasise on parameters for prediction of difficult direct laryngoscopy. The goal of direct laryngoscopy is to get a, at least partial, view of the vocal cords. And the goal for endotracheal intubation is to place a tube through these vocal cords. In order to get a direct view of the vocal cords the alignment of three axes in the head and neck region is necessary. The angle between the oral, pharyngeal and tracheal axis should become zero as visualised in figure 2.1 [13]. The ease or difficulty with which this can be achieved depends on anatomical structures and mobility of these structures. Greenland et al. described a method for direct laryngoscopy and airway assessment based on a three column model correlated with the alignment of the three axes [5,14]. They divide the involved structures in three regions: the anterior, middle and posterior columns. Subsequently, each column is divided in anatomical or functional parameters for direct laryngoscopy. In the following section, we use the nomenclature as described by Greenland et al. [5].



(a) (b) (c)

Figure 2.1: Axes involved during intubation and the necessary alignment of those axes [13].

Anterior column

During laryngoscopy, the anterior column predominantly consists of a number of movements. The mouth is opened, the mandible is anteriorly displaced and the submandibular tissue is anteriorly displaced and partially compressed [13]. The ability to do this depends mainly on three factors. The volume of the submandibular space, the compliance of the submandibular tissues and the mobility of the temporomandibular joint (TMJ).

Volume of submandibular space. Absolute volume of the submandibular space is related to the space necessary for compression of submandibular tissues during laryngoscopy with the laryngoscope. After correct positioning of the patient and inducing the laryngoscope, this is the last step necessary of laryngoscopy to get a view of the vocal cords. Parameters related to absolute submandibular volume are: TMJ-lower incisor distance (correlates to mandibular length), incisor hyoid distance or thyromental distance (TMD), and distance between the left and right TMJ (Figure 2.2). Relative submandibular volume is related to tongue volume relative to oral cavity and prominent top incisors. A relative reduction of volume due to tongue volume/oral cavity ratio causes difficulty with compression of the submandibular tissue during DL. Mallampati score is used as a parameter for this ratio. The Mallampati score is described in the section below. Prominent top incisors reduce intubation sight because the laryngoscope blade needs to be inserted more vertical which is a disadvantage for aligning of the three axes. This can be corrected by a larger anterior displacement of the mandible but with the same anterior displacement this results in a reduced relative submandibular volume.

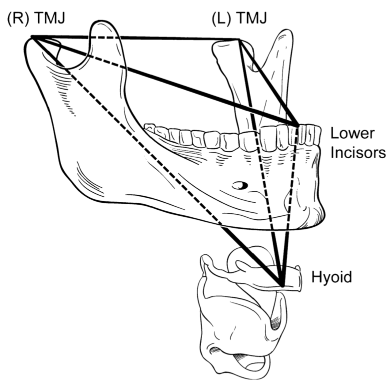


Figure 2.2: Schematic visualisation of the submandibular space [14].

Compliance of the submandibular tissue. Submandibular tissue is partially compressed anteriorly and laterally by the laryngoscope blade. Compliance of the tissue is of importance for the laryngoscopic view. Low compliance result in more difficulty with compression of the tissue and therefore more difficulty for a good view of the vocal cords. Some reasons for low compliance of submandibular tissue include previous radiotherapy, masses, haemorrhage, infections and severe burns in the neck and jaw region. No broadly used test for tissue compliance in the context of difficult intubation is known in literature but low compliance can have a significant effect on the ease of intubation.

Mobility of the temporomandibular joint (TMJ). The last step of the alignment of the oral axis with the pharyngeal and tracheal axes is the anteriorly displacement of the mandible and the displacement and compression of submandibular tissue with the laryngoscopy blade. The mobility of the temporomandibular joint is based on a hinge movement and gliding movement of the maxilla. Limited function of the TMJ results in difficulty with alignment of the three axes. Hinge dysfunction results in a limited mouth opening and sliding dysfunction in the inability to protrude the mandible in front of the maxilla. Therefore, testing of the TMJ function is relatively easy but it can give inconclusive results if dysfunction is related to muscular contraction. Muscular contractions can be averted by a neuro-muscular blocking drug which are of standard use before endotracheal intubation. If the dysfunction is due to calcifications or bony fractures disrupting the joint surface the limited function can complicate intubation.

Middle column
The middle column assessment mainly consists of checking for the passage of the airway in this section. Conditions that could block the passage are laryngeal tumours or lingual tonsillar hypertrophy. Physical examination of the patient for this column is possible with imaging modalities such as: nasopharygoscopy or CT, CBCT and MRI. CT, CBCT and MRI images could be used for virtual laryngoscopy which could be more beneficial, especially for anaesthesiologists who lack experience with the assessment of 2D radiology images. The correlation of these examinations with laryngoscopy is not highly accurate, because patients are awake and not under the influence of muscle relaxation drugs during these examinations. Without a specific reason for one of these examinations, they are not performed and therefore, for most patients, information about the middle column is not available. The most reliable information about the middle column is a recent history of direct laryngoscopy. This is also the most reliable information for the overall prediction of difficult intubation.

Posterior column
The posterior column is mostly related to the alignment of the pharyngeal and tracheal axes which is best achieved by placing the patient in the sniffing position. This section is described at the end but it is the first step and of great importance during preparation of intubation [15]. The ability for placing the patient in this position depends on the ability to flex the cervical spine and extend the occipito-atlanto-axial complex [14]. Several conditions can limit the movement of the cervical spine such as, surgical fusion, rheumatoid arthritis, ankylosing spondylosis or external stabilisations forces. Assessment of the posterior column is focused on the mobility of the neck and it corresponding joints and muscles.

Difficult direct laryngoscopy prediction parameters

In the section below the most commonly used prediction parameters known in literature are summarised. Based on this information a selection was made for the intra-operative prediction parameters used for documentation at the Radboud UMC hospital.

Mallampatie score. The Mallampati test is used to classify the visibility of pharyngeal structures and is first described by Mallampati and later modified by Samsoon and Young [16–18]. The test is carried out with the head in neutral position, full mouth opening and maximum protrusion of the tongue. Based on the pharyngeal structures visible in this position the score is I to IV (Figure 2.3). The Mallampati test is used as a tool to evaluate the relative submandibular volume of the anterior column. Class I-II is related to easy DL and class III-IV is related to difficult DL [19]. Various studies show a poor to moderate discriminative power for difficult laryngoscopy prediction, with an average sensitivity of 50% and specificity of 85% [19–21]. Combining information of prediction studies in a receiver operating characteristic curve (ROC), give an area-under-the-curve (AUC) of 0.82 [20]. This is when only the Mallampati score is used to predict difficult laryngoscopy and thereby difficult intubation. There is a general consensus among anaesthesiologists that Mallampati test alone has insufficient discriminative power for the prediction of difficult intubation. This statement it strengthened by a large meta-analysis of Lundstrøm et al. involving 177,088 patients [22].

Thyromental distance (TMD). The thyromental distance is the distance between the mentum and the thyroid notch in a position with full neck extension. The TMD is used as a tool to evaluate the absolute submandibular volume of the anterior column. This test is a measure for the space that is needed for the translation and compression of the tongue during DL. Sometimes the TMD is measured in fingerbreadths or categorised in a specific set of ranges. Using fingerbreadths is highly inaccurate and using fingerbreadts as a distance significantly decreases the performance of the TMD as predictive test compared to objective measurement [24]. At the Radboud UMC the TMD is categorised in three ranges: Normal >7 cm, limited 6.5-7

(a) (b) (c) (d)

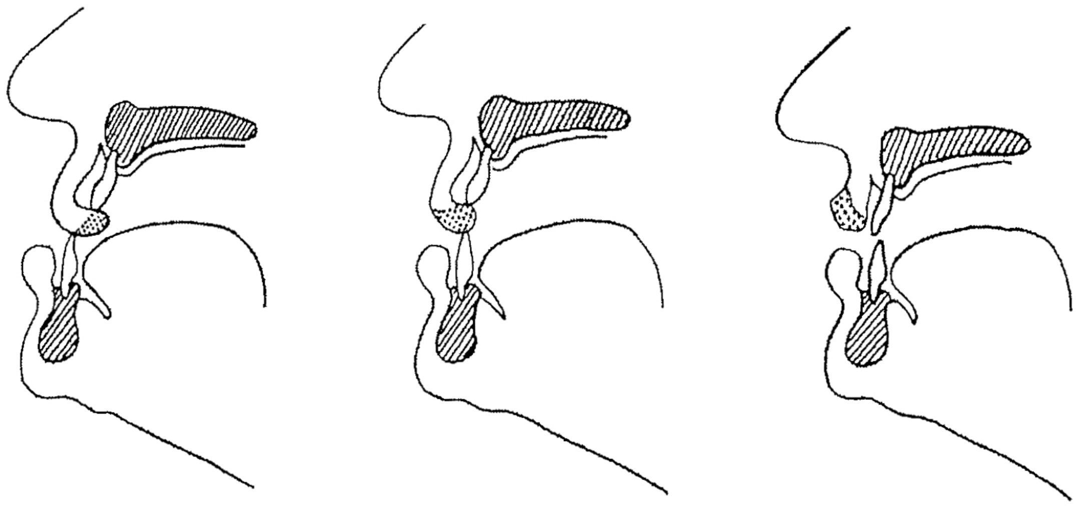
Figure 2.3: Mallampati classification based on the visibility of pharyngeal structures. a: Class I, Soft palate, uvula, fauces, pillars visible. b: Class II, Softpalate, uvula fauces visible. c: Class III, Soft palate, base of uvula visible. d: Class IV, Only hard palate visible [23].

cm or severely limited <6.5 cm. There is no general consensus about the TMD cut-off value from when intubation is difficult, but a smaller distance corresponds to a more difficult DL. A low chosen cut-off value will result in a higher sensitivity and a lower specificity. Cut-off values for prediction performance estimation in literature vary between 7 and 4 cm [20]. Due to this heterogeneity, a wide range of sensitivities and specificities are reported. Combining information of prediction studies in a receiver operating characteristic curve (ROC), give an area-under-the-curve (AUC) of 0.64 and a pooled sensitivity and specificity of 20% and 94%, respectively [20]. Based on the results, the predictive value of this bedside test alone is not good enough for accurate prediction of difficult laryngoscopy.

Figure 2.4: Schematic visualisation of the thyromental distance (TMD) [25].

Inter-incisor gap (mouth opening). Inter-incisor gap is the distance measured between the upper and lower incisors at full mouth opening. This measurement relates to the hinge function of the TMJ, the space available for induction of the laryngoscopy blade and the exposure of the larynx. These functions are all related to the anterior column. As with the TMD, mouth opening also has different cut-off values in different studies but in general a decreased mouth opening is an indication for difficult DL [20]. At the Radboud UMC the inter-incisor gap is categorised in three ranges: Normal >5 cm, limited 3.5-5 cm or severely limited <3.5 cm. Rose et al. used a cut-off value of 2 fingerbreadths and found a relative risk of 10.3 for decreased mouth opening, a sensitivity of 5% and a specificity of 99% [26]. Other studies reported a sensitivity of 40% and 26% and a specificity of 76% and 95% [27, 28]. These result show that the mouth opening alone is not a good discriminating value for difficult DL but that a small mouth opening gives a high risk for difficult DL.

Upper lip bite test (ULBT). The upper lip bite test is an indirect measurement for a combination of the sliding function of the TMJ and prominent top incisors. Therefore, this test is also part of the anterior column. The test is categorised in three classes according to the following criteria: class I = lower incisors can bite the upper lip above the vermilion line (Figure 2.5a); class II = lower incisors can bite the upper lip below the vermilion line (Figure 2.5b); class III: lower incisors cannot bite the upper lip (Figure 2.5c) [29]. Results according the upper lip bite test show a sensitivity of 77%, 55%, 28% and 8% and a specificity of 89%, 97%, 93% and 98% [29–32]. The big difference in sensitivity is probably due to the fact that the earlier study of Khan et al. used only two observers for the ULBT where the study of Eberhart et al. used 25 observers. The study of Myneni et al. reported a very low sensitivity of 8%. The studies with the lowest sensitivity included the most patients, 1425 and 6882. Therefore, there results are the most reliable results. Also, this parameter shows that a negative test is not a guarantee for an easy DL. Again, this parameter alone is a weak discriminator for difficult DL.



(a) (b) (c)

Figure 2.5: Schematic visualisation of the upper lip bite test (ULBT) [29].

Neck extension. The neck consists of many joints and one of the possible movements of the neck is the extension. Neck extension is an important movement for placing the head in a sniffing position. Neck extension give a good prediction about the possibility of placing your head in a sniffing position. The neck extension can be measured in a variety of ways and with varying accuracy. For instance, the extension can be measured in degrees from neutral position to maximal extension and also indirect, as the difference in distance of TMD between neutral and maximal neck extension. All possibilities aim to get an impression about the posterior column involved in intubation, according to the model described at the beginning of this chapter. Pathologies involving the dysfunction of neck mobility, such as ankylosing spondylitis, are known to have a disadvantageous effect on the ease of direct laryngoscopy [33]. Again, with neck extension the measuring technique and the chosen cut-off value have a high influence on the sensitivity and specificity of the parameter. Known sensitivities are 10%, 40%

and 88% and specificities of 98%, 99% and 60% [28, 34, 35]. This shows that this parameter for evaluation of the posterior column alone is not a good predictor for difficult DL.

Obesity and Neck circumference (NC). Obesity is a risk factor for difficulty at tracheal intubation [2]. Obesity is standard monitored in the patient's electronic medical record by body mass index (BMI). At Radboud UMC, BMI is not used as a predictor of difficult intubation. Neck size is a parameter which is used and therefore creating the expectation of a more local measurement of obesity. However, the neck circumference does not differentiate between the ratio of muscle and fat. Research has shown that neck circumference is a good predictor for difficult intubation for obese patients, sensitivity of 92% and specificity of 84% [36]. Where BMI alone had a sensitivity of 83% and a specificity of 50%. As stated above BMI is an independent predictor for difficult intubation and neck circumference an even better predictor. Further research showed a neck circumference to thyromental distance ratio (NC/TMD) in obese patients as an even better predictor, sensitivity of 88% and a specificity of 83% [37]. The odds ratio for NC/TMD was three times higher compared to the NC alone in the same study. Therefore, neck size is kept as a parameter for prediction. In a computer model the performance of this parameter could be increased when combined with the BMI or thyromental distances as described in literature.

Multivariate models. Various research including this short summary has shown that no single prediction parameters has a high discriminative value. Lots of parameters are known as a risk factor for difficult intubation but all lack the value of being a good discriminator. Therefore, a combination of risk factors seems to be the only way of increasing the predictability of difficult direct laryngoscopy. There is a limit to the number of parameters because of the added complexity for each extra parameter. Also, Nørskov et al. showed us that just adding extra parameters does not increase the correct prediction of difficult intubation [38]. Lots of multivariate research is done to find the perfect combination of parameters and the formula to combine the values [28, 39–42]. One multivariate predictor test is standardised and is known as the Wilson Risk Score [42]. The advantage of this test is the higher homogeneity in test results in varying studies whit a pooled sensitivity of 46% and specificity of 89% in a total of 5 studies [20]. In the study of El-Ganzouri et al. involving 10.507 patients and using seven predictive parameters a sensitivity of 65% and specificity of 94% was achieved [28]. Later study by Langeron at al. in 2012 used a computer-assist model with nine parameters resulting in an area under the receiver operating characteristic curve of 0.86 [43]. This is a significant improvement compared to the model used by Naguib et al. and a using a simple logistic model on the same data [41]. A big advantage for multivariate models is more choice in determination of a cut-off value for the different models. With known characteristics of the model it is possible to make a choice between a better sensitivity and a lower specificity or the other way around.

Chosen parameters for pre-operative documentation

Based on the literature study above and the already documented parameters in the clinic, a selection was made for the documented prediction parameters. Mallampati grading, mouth opening and neck mobility were the most commonly performed bedside interactive tests [2]. The results show that the Mallampati, mouth opening, neck mobility/extension, TMD and neck circumference are the best univariate predictors for difficult direct laryngoscopy. Therefore, we made the decision to use these parameters for documentation during the pre-operative consult. Age, sex, BMI, length and weight are possible useful parameters for calculation of ratios and evaluation of study group. These are already standard documented parameters so no extra attention was needed for these parameters.

Besides the prediction parameters, a statement of the anaesthesiologist about the expected difficulty is added as required input. After reporting the outcome of the prediction parameters, a dialog box needs to be checked if the airway is expected to be difficult or not. And if the expectation is difficult, there are five options to check to specify the expected difficulty. The options are: difficult mask ventilation, laryngeal mask placement, laryngoscopy, endotracheal intubation or surgical airway. The options are useful so the data can also be used for further research.

2.2.2 Intra-operative parameters

Studies about difficult airway require some kind of discrimination between the easy and difficult group. Literature is not conclusive about this discrimination and a lot of studies use their own way to determine the groups. In this paragraph, we studied the literature about the possible ways to divide the patients' groups and therefore the different known definitions of difficult airway, difficult intubation and difficult direct laryngoscopy.

Airway definitions

Difficult direct laryngoscopy. Difficult laryngoscopy, direct or indirect, depends on the laryngeal exposure using direct laryngoscopy. Laryngeal exposure is generally quantified using the Cormack-Lehane grade [7]. Most authorities such as: the American society of anaesthesiologists task force, Canadian airway focus group and the British royal college of anaesthetists and the difficult airway society, all define grade 1 and 2 as easy, and 3 and 4 as difficult laryngoscopies [2, 15, 44, 45]. Definition of difficult laryngoscopy is the parameter with the most consensus. Difficult laryngoscopy is used a lot to define difficult intubation because there is a relation between the parameters [2, 6]. However, difficult intubation is not always the result of difficult laryngoscopy. Intubation can be difficult even if there is a Cormack-Lehane grade 1 or 2 visibility of the vocal cords. Therefore, difficult intubation is a more extended definition.

Difficult intubation. As stated above, difficult intubation is more than difficult laryngoscopy alone. Difficult intubation is sometimes defined as difficult laryngoscopy, but mostly the difficult intubation definition is based on more parameters than difficult laryngoscopy alone. The extensiveness of parameters causes a lack of consensus about the definition of difficult intubation. Each study uses her own definition and parameters. In an attempt to reach consensus, the intubation difficulty scale (IDS) is introduced by Adnet et al. [46]. Unfortunately, the scale is not used in most studies [46]. However, many studies use a subset of parameters for this scale to define difficult intubation. The parameters of the scale are: Number of attempts, number of operators, number of alternative techniques, Cormack-Lehane grade, lifting force required (normal or increased), Laryngeal pressure used, Vocal cord mobility. The score gives values to all parameters and a total of more than 5 is seen as a difficult intubation. Law et al. defines difficult intubation as one or all of the following: multiple attempts or more than one operator; use of an adjunct tool such as a tracheal tube introducer; An alternative intubation device is required after unsuccessful use of primary plan [44]. Apfelbaum et al. uses: "Tracheal intubation requires multiple attempts, in the presence or absence of tracheal pathology" [45]. Caplan et al. uses measurements of time and attempts. They label an intubation difficult if intubation requires more than three attempts or if intubation requires more than ten minutes [15]. A large and recent Danish study form Nørskov at al. used the measurements similar of Law et al. where intubation is scored difficult if more than two attempts were necessary or other equipment than only a conventional laryngoscope was used [38].

Difficult airway. Definition according to Law et al.: "an airway where an experienced provider anticipates or encounters difficulty with any or all of face mask ventilation, direct or indirect (e.g., video) laryngoscopy, tracheal intubation, supraglottic airway device (SGD or SAD) use, or surgical airway" [44]. Based on this definition it can be stated that a difficult airway is a combination of difficulty on one of all possible ways to ventilate a patient's lungs.

Chosen parameters for intra-operative documentation

Based on the literature study above and the already documented parameters in the clinic, a selection was made for the documented intra-operative parameters. Documented parameters are: number of attempts necessary, number of experienced operators, change of primary plan (yes or no), Cormack & Lehane grade and the use of extra or change of materials or technique. With these parameters it is possible to perform our baseline study and replicate most of the definitions, thereby create an easy and difficult intubation group according to the same values as other studies.

2.3 Discussion

This study was performed to adjust pre- and intra-operative documented parameters for enabling follow-up studies about difficult laryngoscopy, difficult intubation and difficult airway. We did research about anatomy involved in intubation, prediction parameters for difficult laryngoscopy and intubation, and the use of different definitions of difficult laryngoscopy, intubation and airway. Eventually we compared the study results with the used parameters in the clinic and we change the requested documentation in the patient's electronic medical record if necessary.

Prediction parameters were chosen with knowledge of the theory of direct laryngoscopy based on the three column approach. Parameters were chosen in such way that all possible columns were included and that all possible movements and or volumes were included by a parameter as well. The only parameter that was not included, is the upper lip bite test for evaluating the sliding function of the TMJ. This parameter was not used because the results were inconclusive and it was not already used by the anaesthesiology department. We decided that the results were not conclusive enough to give the department extra training and information about using this bedside test. The middle column is not possible to evaluate with a simple bedside test, so information of this column is only available if its already known by the patient or other departments and documented in the patient's medical record. If there is relevant information about the middle column this is added as a note together with the standard pre-operative airway parameters.

At the pre-operative documented information, a question was added to give a final verdict if the airway is predicted difficult or not. If the airway is predicted difficult the anaesthesiologist has to document what is predicted difficult: difficult mask ventilation, laryngeal mask placement, laryngoscopy, endotracheal intubation and/or surgical airway. This information is important for comparing prediction performance between the anaesthesiologist and a possible new prediction model. When the changes are applied to the workflow of difficult airway prediction the information can be used to monitor changes in performance.

Intra-operative more changes were made. Before only the Cormack-Lehane grading was documented. Number of attempts, number of operators, used extra material or techniques and change of plan were added. This information will be used to divide patients in a group of easy intubation and a group for difficult intubation.

The downside of this study is the comparison of parameters based on their sensitivity and specificity and the usage of different cut-off values for calculation of these performance values. Most prediction parameters have a cut-off value to differentiate between easy or difficult to intubate. For instance, TMD smaller than 6 cm is labelled as difficult for intubation. Based on this cut-off value the sensitivity and specificity are calculated. These outcomes vary highly with the chosen cut-off value. A smaller value results in a smaller sensitivity and a higher specificity. Because studies use different cut-off values it is hard to compare results of different studies. Besides the difference in cut-off values, also the low incidence and the use of specificity for difficult intubation is a problem. Low incidence like difficult intubation and a prediction parameter results almost automatically in a high specificity. For instance, if the incidence of difficult intubation in a large group is 2% and we would predict an easy intubation for everyone in the group, this would result in only a 2% miss of false negatives and thus a specificity of 98%. This seems to be a very high specificity but with such a large difference in group size between easy and difficult to intubate it is fairly easy to reach this specificity. Depending on the incidence of difficult intubation the value of the specificity must be critically examined. Due to differences in incidence between studies it is harder to compare results. However, we used sensitivity and specificity in our study to compare prediction parameters because these performance values are used in general. Therefore, we could compare more studies with each other.

The literature study resulted in changes in the digital documentation system at the Radboud UMC. Pre-operative documented parameters for the prediction of a difficult airway were not changed because al parameters with the highest possible sensitivity and specificity were already standard for documentation. Already documented pre-operative parameters were: Mallampati score, TMD, mouth opening, neck extension and neck circumference. The only addition to the pre-operative documentation included a final verdict of the anaesthesiologist about the predicted difficulty, yes or no. This is necessary for a baseline study to calculate the performance of the current prediction method and for monitoring possible changes in the method in the future. Intra-operative changes were made in the documentation of the used airway management methods. Already documented parameters about intubation were: Cormack-Lehane score and used materials. Added mandatory fields for documentation were: number of attempts, number of operators, use of extra materials or techniques and the change of primary airway plan. The additional parameters will be used for the discrimination between easy or difficult intubation. The definition of easy and difficult intubation varies between studies and mainly is a specific combination and scores of these parameters. With the addition of these parameter we can divide our patients in the same way as most other studies in literature do.

Chapter 3

DTM landmarking study

Various research including the summary in chapter 2 has shown that no single prediction parameter has a high discriminative value. Lots of parameters are known as a risk factor for difficult intubation but all lack the value of being a good discriminator. Therefore, a combination of risk factors seems to be the only way of increasing the predictability of difficult direct laryngoscopy. There is a limit to the number of parameters because of the added complexity for each extra parameter. Also, Nørskov et al. showed us that just adding extra parameters does not increase the correct prediction of difficult intubation [38]. Lots of multivariate research is done to find the perfect combination of parameters and the formula to combine the values [28, 39–42]. In 2012 Langeron et al. used a computer-assisted model with nine parameters resulting in an area under the receiver operating characteristic curve of 0.86 [43]. This is a significant improvement compared to the model used by Naguib et al. and the use of a simple logistic model on the same data [41]. The improvement of results with the use of a computer-assisted model proves the complexity of the problem and the usefulness of machine learning for a better prediction of difficult intubation. A big advantage for multivariate models is more choice in determination of a cut-off value for the different models. With known characteristics of the model it is possible to make a choice between a better sensitivity and a lower specificity, or the other way around.

Recent studies showed that analysis of parameters derived from facial photographs are useful for the prediction of difficult intubation [1, 9]. These studies used 2D photographs of patients. One study used only the natural head pose and the other study used also positions to measure mouth opening and neck extension. Based on these studies the Radboud UMC started a pilot study to make a computer model based on 3D facial stereophotographs with different head positions. The hypothesis was that 3D stereophotographs could improve accuracy because there is a possibility to measure absolute values like distances, instead of only ratio's and angles. This pilot study contained too little subjects to draw solid conclusions about the accuracy of the method but it showed the feasibility and potentials of the method. In this study, this path was extended with an automatic analysis method for the 3D stereophotographs. The method was developed for the automatic annotation of landmarks on 3D stereophotographs. Later, these landmarks can be used to calculate features which are used as input for machine learning. The landmarks and facial postures necessary for the development of an accurate prediction model are not yet fully known and therefore, a highly adaptive landmarking algorithm is preferred. In this study, we developed such landmarking algorithm and studied the performance of this algorithm. The method was also compared to two commercially available landmark annotation systems. The algorithm is based on deformable template matching (DTM) and will be further referred to as DTM.

The research question for this path was: What is the performance of a new developed deformable template matching landmark detection algorithm for use with 3D stereophotographs?

3.1 Methods and Materials

3.1.1 Study cohort

The developed landmarking algorithm was verified on 3D stereophotographs of 20 healthy volunteers. 10 males and 10 females, all adults. The landmarking algorithm was tested against commercially available landmarking algorithms in 3dMDvultus (3dMD, Atlanta, GA, United States) and Vectra (Canfield, Fairfield, NJ, United States). Each system detected a set of landmarks. Corresponding landmarks between a commercial system and the DTM method were compared with each other. The commercial landmarks systems performed better if the 3D stereophotographs were made on a system of the same manufacturer. Therefore, two stereophotographs were taken from each volunteer. One on 3dMD face system and one on a Canfield VectraXT face system, giving a total of 40 photographs. Stereophotographs were taken in natural head position. Scalp hair covering the face was cleared away through the application of a hairnet or hairband. At both jaw angles and the thyroid notch green stickers were placed for automatic detection of these landmarks. These landmarks could be of added value for the difficult airway prediction model and are without manual annotation, hard to located on 3D stereophotographs. These landmarks are easy to palpate. In table 3.1 the landmarks detected by the different systems are shown.

Table 3.1: Detected landmarks by different systems.

Landmark	3dMDvultus	Vectra	DTM
Exocanthion right (ex_r)	✓	✓	✓
Endocanthion right (en_r)	✓	✓	✓
Endcanthion left (en_l)	✓	✓	✓
Exocanthion left (ex_l)	✓	✓	✓
Nasion (n)	✓	✓	✓
Pronasale (prn)	✓	✓	✓
Subnasale (sn)	✓	✓	✓
Alare right (al_r)	✓	✓	✓
Alare left (al_l)	✓	✓	✓
Cheilion right (ch_r)	✓	✓	✓
Cheilion left (ch_l)	✓	✓	✓
Sublabiale (sl)	✓	X	✓
Pogonion (po)	✓	✓	✓
Zygion right (zy_r)	X	X	✓
Zygion left (zy_l)	X	X	✓
Glabella (g)	✓	✓	✓
Jaw angle right green marker (ja_r)	X	X	✓
Jaw angle left green marker (ja_l)	X	X	✓
Thyroid notch green marker (thy)	X	X	✓

3.1.2 Deformable template matching landmark algorithm

Commercially available landmark systems are bounded by a standard set of landmarks and head positions. To be able to use the system for the analysis of 3D stereophotographs made for the prediction model for the anaesthesiology department, the ability to add some extra functionality is needed. At the moment, it is not know exactly which landmarks are necessary for creating the best possible model. Therefore, the ability to delete or add landmarks is preferable. Secondly, the 3D stereophotographs for the model were made in three different patient positions and postures. The commercial algorithms normally are only designed for natural head position with natural face pose. Finally, the algorithm must be able to find the manually annotated landmarks (green stickers) automatically. Therefore, the decision was made to develop our own landmark algorithm. The landmark algorithm was completely developed in programming software package MATLAB version 9.1.0 (MathWorks Inc., Natick, MA, United States). The section below gives a detailed explanation about the structure and functionality of the sub functions of the complete algorithm.

General principal

The algorithm consists of nine different blocks or sub methods (Figure 3.1). Apart from block one and two (Initialisation and Patient(s) selection), each block can be looped according to the number of patients selected.

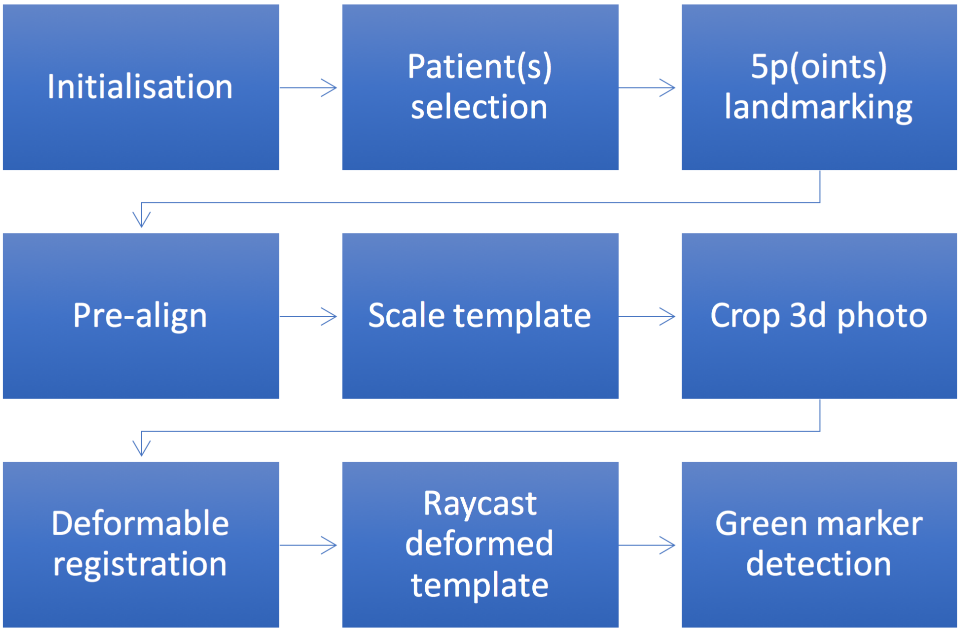


Figure 3.1: Flow chart of the deformable template matching (DTM) algorithm with all sub functions.

During the initialisation method, path allocation for sub programs and patient files was performed and all parameters for the next eight blocks were chosen and stored in one main parameters structure. In the next block (patient(s) selection), .obj files for the landmark algorithm were selected in a pop-up window. The 3D stereophotograph's mesh, texture file (.bmp or .jpg) and link between those two were loaded. The information was saved in a structure with the same categorisation for each patient. Also, empty categories were allocated for further use during the rest of the algorithm.

At this moment, all information for the sub functions was available in the structures *Patients* and *Parameters*. Therefore, the setup of al sub functions could be made the same and has the following lay-out: $[Patients] = functionname(Patients, Parameters)$. Here, the function uses the two structures as input and the updated patients structure as output.

The algorithm makes use of a facial template with all desired landmarks allocated. The face template is an average face made from seventy 3D stereophotographs of healthy randomly selected adults. Autodesk MeshMixer (version qt 4.6.8, San Rafael) was used to manually segment this average face into 29 facial esthetic units. The facial esthetic units indicated on the average face were checked by two different head and neck surgeons. The template with the facial esthetic units is displayed in figure 3.2. This template was matched with a deformable image registration algorithm to the patients 3D stereophotograph. A non-rigid coherent point drift (CPD) algorithm was used for the deformable registration [47]. After the registration, the position of the landmarks of the deformed template were projected on the 3D stereophotograph. Not only landmarks but all known positions on the template can be adopted on the patient's stereophotograph, such as the facial esthetic units.

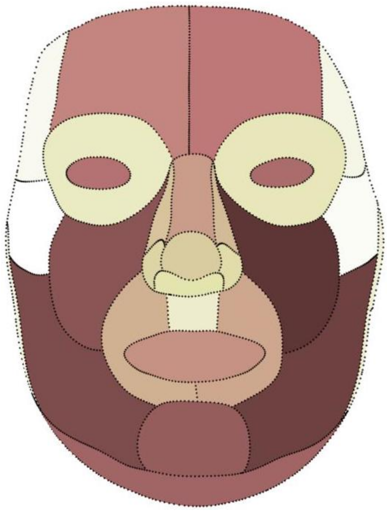


Figure 3.2: Facial template with facial esthetic units.

A set of pre-processing steps were necessary to facilitate a good performance of the deformable registration. These include a pre-alignment and cropping of the 3D stereophotograph such that the boundaries are approximately alike the template. Some functions were designed in such way that it is possible to enable parallel processing. This saves time when more than one patient was selected. A more detailed explanation of the sub functions is given in the subsections below.

5p(oints) landmarking

This sub function was used to automatically detect five (left and right pupil, nose tip and left and right mouth corners) facial landmarks. These landmarks were used to align the patient's 3D stereophotograph in a standard reference frame. The landmarks were also used to scale the template. The scaled template was used to crop the 3D stereophotograph to enable it as input for the CPD algorithm. All 3D stereophotographs were stored as .obj file format with an .bmp or .jpg texture file. This texture file contained, depending on the number of pods used during the photo acquisition, several 2D photos of the face in one picture (figure 3.3a and 3.3b). These texture files were used to give the 3D mesh a texture/color. Therefore, the .obj format contains linking data between coordinates in the texture file and the vertices of a 3D mesh. The 5p landmarking function uses an open source 2D facial landmark algorithm to detect five facial landmarks on an 2D image. With the texture link these landmarks were related to their position on the 3D mesh.

The algorithm starts with a face detection of the texture file. The face detection and landmark algorithm make use of pre-trained deep learning models. The models were developed and trained in deep learning framework Caffe and used in MATLAB with a special interface [48]. The algorithm of face detection and landmark detection uses a cascade of convolutional neural networks (CNN) [49]. The face detection was performed in three steps in which the last step

(a) 3dMD texture file

(b) Vectra texture file

(c) 3dMD face and pose detection

(d) Vectra face and pose detection

(e) 3D stereophotograph right view

(f) 3D stereophotograph left view

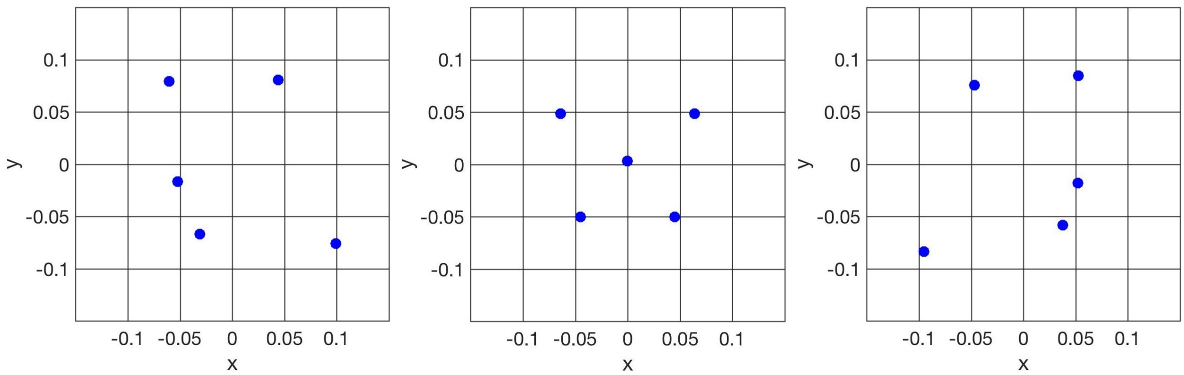
Figure 3.3: 5p(oints) landmarking algorithm.

also detected the five landmarks. The parameters for face detection such as maximal face size and threshold values for each step were optimised for the detection of faces in our texture maps.

Each detected face contained five landmarks. The landmarks are accurate on a frontal image of a face, like on the texture file of a 3D stereophotograph of the VectrXT system (Figure 3.3b). If the face images were acquired from the sides (like on a 3dMD texture image figure 3.3a), it was possible that half of the landmarks were inaccurate (the landmarks on the other side of the face then of which the photograph was taken). In this case the accurate landmarks from both faces were combined. Therefore, a viewpoint detection was developed to find out if a detected face was a frontal picture, a picture from the right side or a picture from the left side. The viewpoint detection used three 2D reference point clouds, one for each viewpoint (frontal, right and left) (Figure 3.4). The reference point clouds represented the general position of the five detected landmarks for each viewpoint. A Procrustes algorithm was performed between the point clouds of all detected faces and all viewpoint reference point clouds. A good fit between a detected face point cloud and a viewpoint reference point cloud means that the detected face viewpoint corresponds with the viewpoint of the reference point cloud. Procrustes calculates a best possible fit of the 2D coordinates of the detected landmarks $\mathbf{y}_n$ to the 2D coordinates of a viewpoint reference point cloud $\mathbf{x}_n$ [50]. Transformation of $\mathbf{y}_n$, with Procrustes results in $\mathbf{y}'_n$. The formulas of the Procrustes algorithm are further explained in appendix B. A goodness-of-fit value was calculated by the algorithm for evaluation. In this case, the sum of squares error d is calculated. This measure was standardised by a measure of the scale of $\mathbf{x}_n$.

$$d = \frac{\sum_{n=1}^{N} \|\mathbf{x}_n - \mathbf{y}'_n\|^2}{\sum_{n=1}^{N} \| \mathbf{x}_n - \bar{\mathbf{x}}_n\|^2} \tag{3.1}$$

Here, $\bar{\mathbf{x}} = \frac{1}{N}\sum_{n=1}^{N} \mathbf{x}_n$ is the centroid of the five 2D coordinates of a viewpoint reference point cloud ($N = 5$). For each detected face, the viewpoint reference point cloud with the lowest d indicates the correct viewpoint. Results of a 3dMD and VectraXT texture images are shown in figure 3.3c and 3.3d.



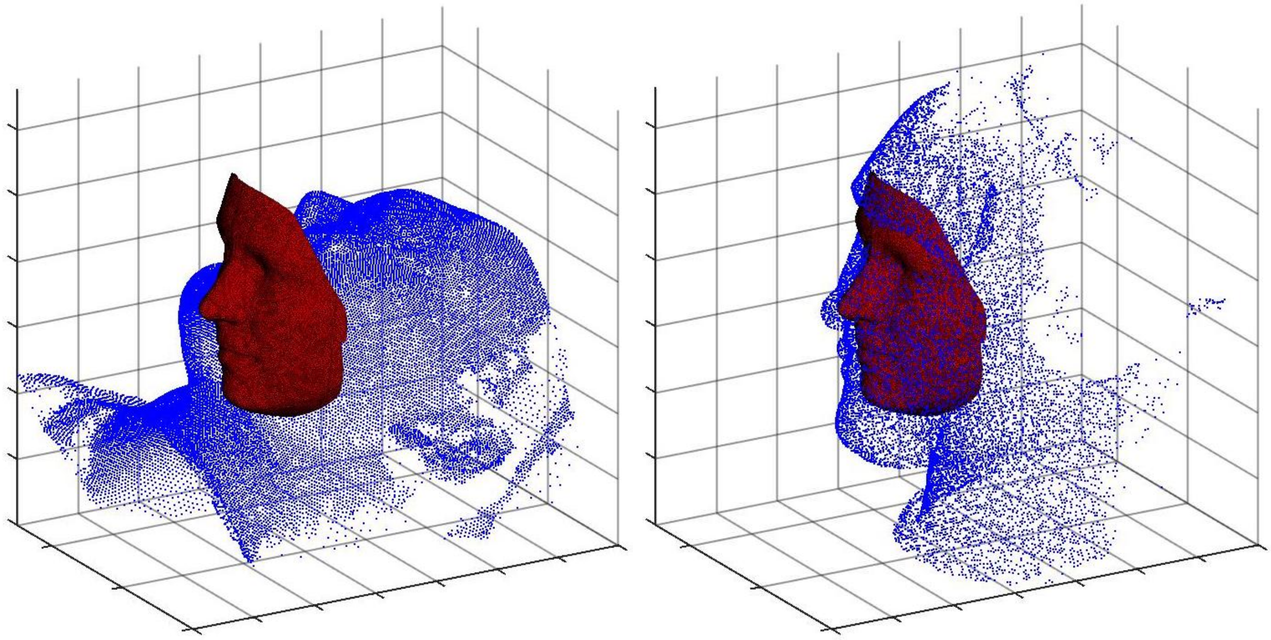
(a) Viewpoint reference point cloud left (b) Viewpoint reference point cloud frontal (c) Viewpoint reference point cloud right

Figure 3.4: 5p(oints) viewpoint reference point clouds.

If a frontal face was detected the five 2D landmarks are used for 3D. If a right and left face were detected, the right pupil and mouth corner from the right face were used and the left pupil and mouth corner form the left face were used. The nose point from both faces were used and the averaged of the points in 3D closest to a mesh vertex becomes the nose point in 3D (Figure 3.3e and 3.3f).

Pre-align

The pre-alignment function aligns all imported 3D stereophotographs to the same coordinate system. The pre-alignment makes use of the five automatically marked landmarks from the 5p landmark algorithm and the same five landmarks manually placed on the face template. A Procrustes analysis was performed to calculate rotation matrix and the translation vector to transform the 3D stereophotograph to the same orientation and position as the template. The formulas of the Procrustes algorithm are further explained in appendix B. In this case $\mathbf{x}_n$ are the five points of the template and $\mathbf{y}_n$ are the five points of the 3D stereophotograph. The rotation matrix is calculated with no reflection allowed and no scaling is used for the pre-alignment of the 3D stereophotograph. In figure 3.5 the effect of the function is visualised.

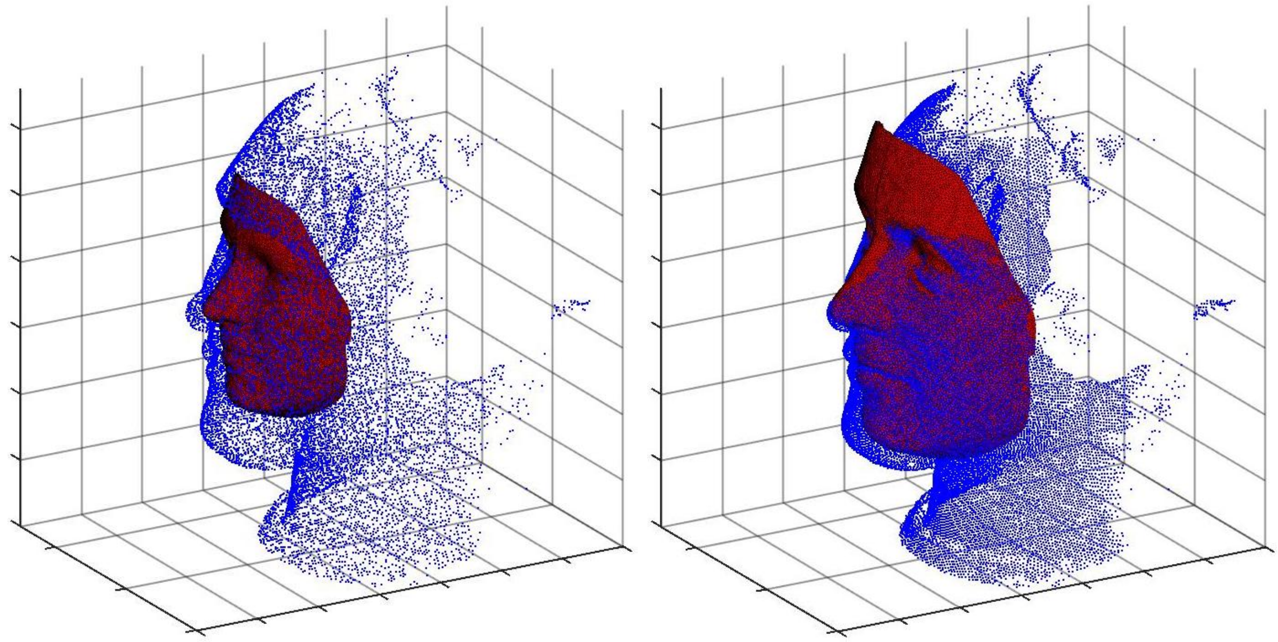


(a) Before pre-align function (b) After pre-align function.

Figure 3.5: Before and after pre-align function. Red = template, Blue = 3D stereophotograph. Equal aspect ratio for all axis.

Scale template

Scaling of the template was necessary to crop the 3D stereophotograph in approximately the same shape as the template. This was required for the deformable image registration algorithm to work correctly on the edges of the stereophotograph. The algorithm works in a similar way as the pre-alignment. This time Procrustes was used to calculate the rotation matrix, scaling component and the translation vector for the transformation of the template to the pre-aligned 3D stereophotograph. The formulas of the Procrustes algorithm are further explained in appendix B. The five landmarks of the stereophotograph and the template were used as input for the Procrustes analysis. In this case $\mathbf{x}_n$ are the pre-aligned five points of the 3D photo and $\mathbf{y}_n$ are the five points of the template. The template was scaled and aligned with the pre-aligned stereophotograph. This time the scaling factor S was used and reflection was not allowed. In figure 3.6 the start and end shapes and position of the template before and after this function are shown.



(a) Before template scaling function (b) After template scaling function.

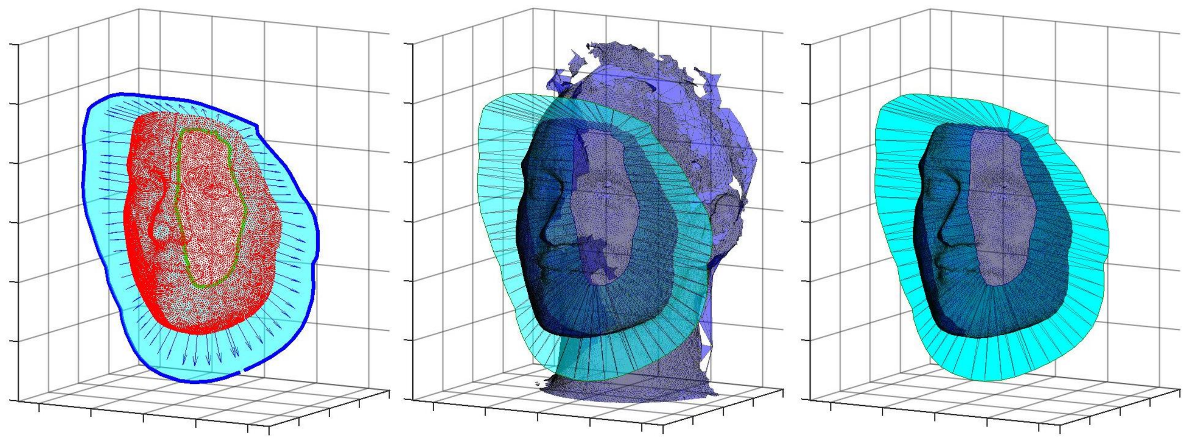
Figure 3.6: Before and after template scaling function. Red = template, Blue = 3D stereophotograph. Equal aspect ratio for all axis.

Crop 3D stereophotograph

As explained before, cropping of the 3D stereophotograph was necessary for a proper functioning of the deformable registration. The registration algorithm works best if the 3D stereophotograph and the template were similar in content. Therefore, the template was used to crop the 3D stereophotograph to create similar edges. The scaled template is the best possible fit of the template with the 3D stereophotograph, based on the five landmarks. The vertex normals on the edges of the template were calculated. Two sets of extra vertices are made, one set is the translation of the edge vertices in the direction of the vertex normal, and the other set in the opposite direction of the vertex normals (Figure 3.7a). The new points set represent an inner ring and outer ring. Between these two rings a mesh was calculated. This mesh intersects the 3D stereophotograph and was used as crop surface (Figure 3.7b). The result was a cropped 3D stereophotograph with similar boundaries as the template (figure 3.7c).

Deformable registration

The landmarks on the scaled face template cannot be transferred to the 3D stereophotograph in the current phase, because the match based on 5 landmarks was not accurate enough (Figure 3.8a and 3.8b). To obtain an accurate match between the template and the 3D stereophotograph, a deformable registration step was used. The matching was performed using a coherent point drift (CPD) algorithm designed by Myronenko et al. [47]. This algorithm is a non-rigid point set registration. CPD uses a maximum likelihood estimation with motion coherence constraint over the velocity field. The CPD algorithm is an iterative process and was used with 300 iterations. Therefore, the face template was able to warp to approximately the shape of the 3D stereophotograph (Figure 3.8c and 3.8d).



(a) Template with edge vertex normals and crop surface.
(b) 3D stereophotograph with crop surface.
(c) Result after crop.

Figure 3.7: Crop 3D stereophotograph function. Red = template, Cyan = crop surface, Blue = 3D stereophotograph.

Raycast deformed template

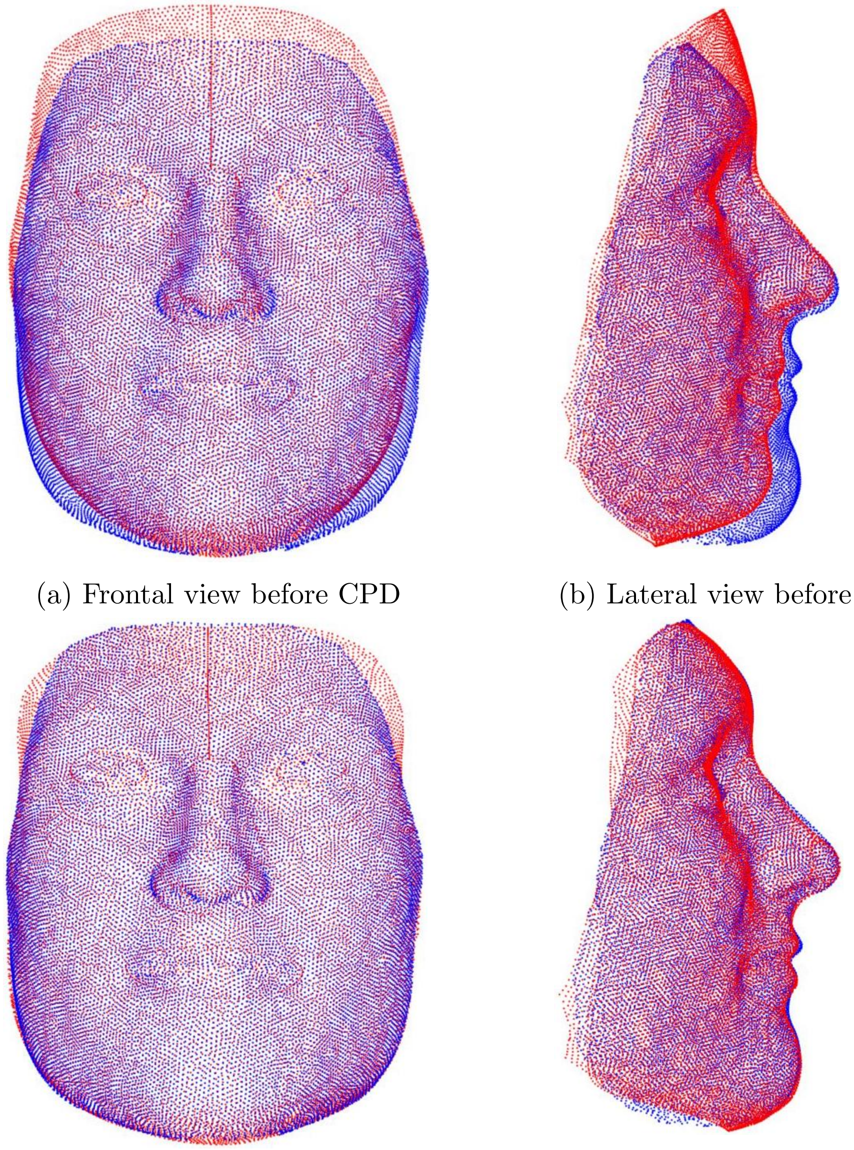
The deformed template approximates the 3D stereophotograph. Known landmarks on the template are linked to specific vertices of the template. To find a point on the 3D stereophotograph that corresponds with a specific point on the template, ray casting was used. Ray casting is the calculation of an intersection of a ray with a surface. In this problem, the ray cast algorithm "shoots" a virtual ray from a vertex on the deformed template in the direction of the vertex normal or in the opposite direction of the vertex normal. Somewhere on the path of the ray it intersects the surface of the 3D stereophotograph. The intersection point between the 3D stereophotograph and the ray of the vertex, is the position of the template's vertex on the 3D stereophotograph. The algorithm loops through all vertices of the deformed template which result in a new array of vertices (Figure 3.9). The order of vertices of this new array is the same as the vertices on the template. The known landmarks on the template correspond to an index of the vertices. Therefore, the landmarks on the 3D stereophotograph are known by using the index of a landmark on the template. Also, the index of vertices defining facial regions of the template are known and can be visualised on the 3D surface (Figure 3.10a and 3.10b).

Green marker detection

Some important facial landmarks who are difficult to detect on only the 3D stereophotograph were marked manually with a green sticker. These landmarks are the thyroid notch for calculation of the thyromental distance (TMD) and the left and right angle of the mandible.

The function uses the texture image for the localisation of the green markers and starts with the calculation of two difference maps. The first difference map is the value of the green channel minus the red channel of the RGB image. The second difference map is the value of the green values minus the blue values. Finally, these two maps were added together. The bigger the value of this final difference map the greener a pixel was on an RGB image (Figure 3.11a).

On the difference map the four green markers were bright (white) spots. To segment these four spots and locate their centres a logical (black-white) image was made. In this image, only the pixels containing the highest one percent values of the image were white (figure 3.11b). However, this image still contains white pixels that were not part of the green markers. A few morphological image operations were used to segment only the four markers. Firstly, an opening operation was performed with a disk-shaped kernel. This operation removes noise because the



(a) Frontal view before CPD

(b) Lateral view before CPD

(c) Frontal view after CPD

(d) Lateral view after CPD

Figure 3.8: Deformable registration with a non-rigid coherent point drift (CPD) algorithm. Blue = cropped 3D stereophotograph, Red = scaled template.

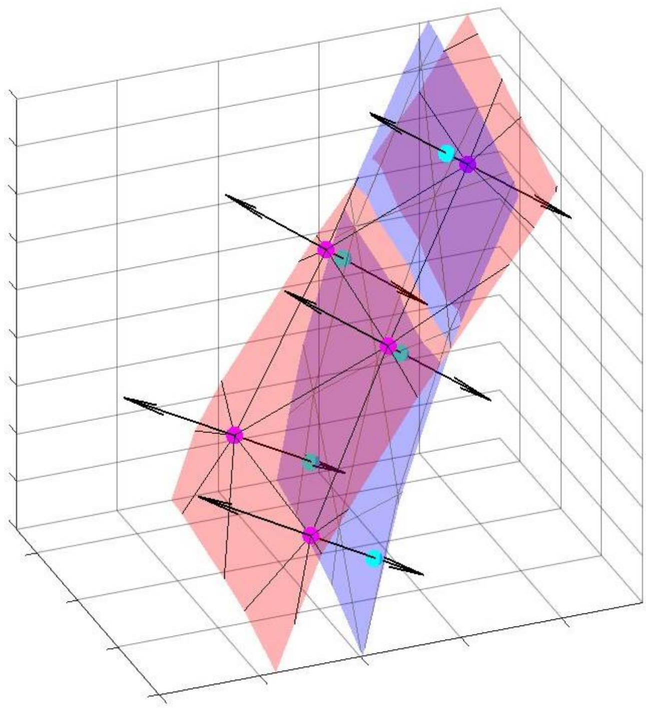
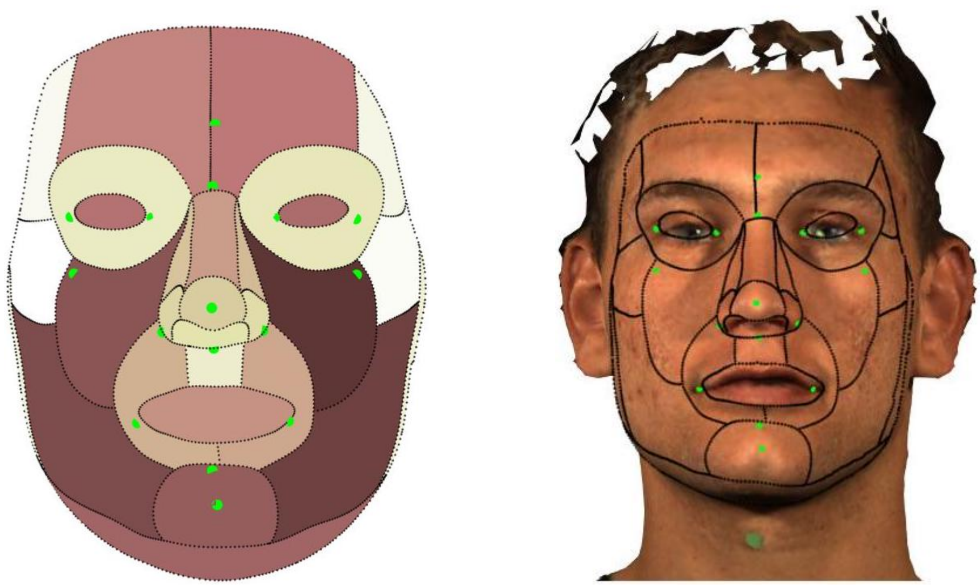


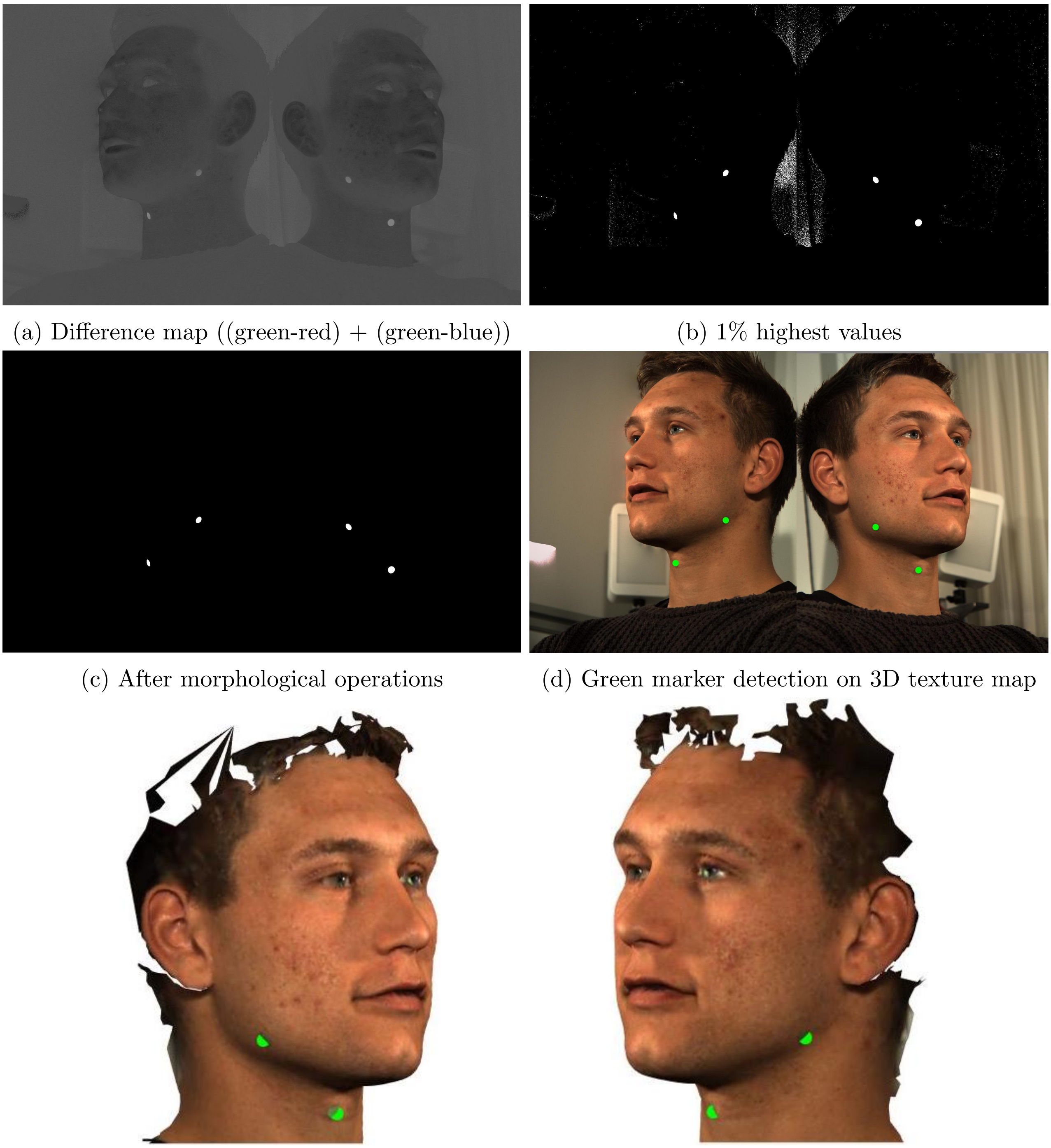
Figure 3.9: Visualization of Raycast function. Red = deformed template surface, Blue = 3D stereophotograph surface, Pink = vertices of deformed template, Blackarrows = rays from the deformed template's vertices (pink) in the direction of the vertex's normals and in the opposite direction of the vertex's normals. Cyan = intersection points of rays with the surface of the 3D stereophotograph (blue).



(a) Template with known facial regions and landmarks

(b) Facial regions and landmarks projected on the original 3D stereophotograph

Figure 3.10: Projection of landmarks and facial regions from template to 3D stereophotograph.



(a) Difference map ((green-red) + (green-blue))

(b) 1% highest values

(c) After morphological operations

(d) Green marker detection on 3D texture map

(e) Result right view

(f) Result left view

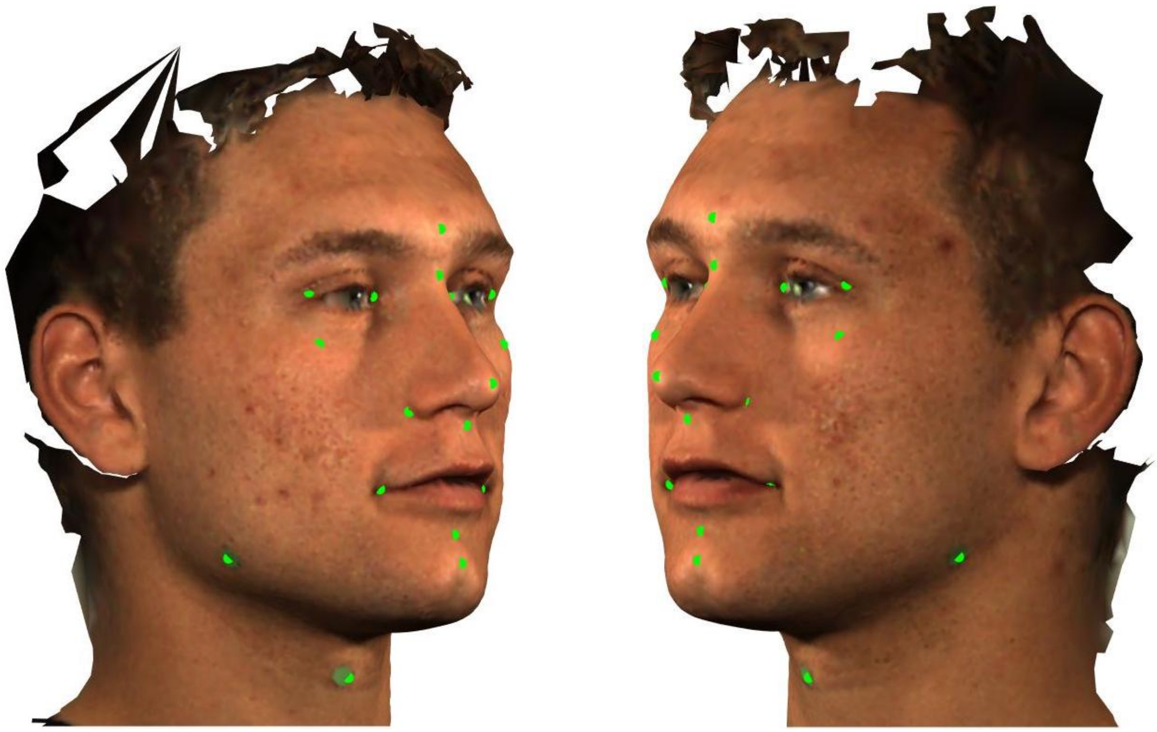
Figure 3.11: Green marker detection algorithm.

noise is mostly some single or small group of isolated white pixels. The four markers were relatively large groups of white pixels. The second operation selected only the four largest groups of pixels. The last operation was an region filling operation and fills all possible little holes in the four isolated groups of pixels (Figure 3.11c). The centres of these four groups of pixels were located by calculating the mean of all pixels in one group. These centres were the positions of the green markers in the 2D texture image (figure 3.11d).

The transfer from the 2D coordinates to 3D coordinates on the surface of the 3D stereophotograph was performed in the same way as done in the 5p landmarks function. The thyromental marker was shown on the left and right image on the texture map of a 3dMD stereophotograph and therefore the markers a both transferred to 3D an averaged to one 3D coordinate (Figure 3.11e and 3.11f).

Final result

The green marker landmarks are added to the already automatically detected 16 facial landmarks to give the final result (Figure 3.12a and 3.12b).



(a) Right view (b) Left view

Figure 3.12: Final result of 19 landmarks automatically annotated on a 3D stereophotograph.

3.1.3 Verification of the algorithm

Performance parameters

Landmark performance can be evaluated in two different ways: (i) golden standard based localisation error; (ii) task-oriented performance [51]. In our study, we do not have a specific task for the landmarking and therefore the impact of the landmarking accuracy cannot be measured. The golden standard (ground-truth) based localisation error was used to evaluate the performance of our method. In this study three experienced observers placed the 16 facial landmarks independently based upon the manual landmarking method described by Plooij et al. (i.e., with the use of a reference frame to ensure optimal accuracy of the landmark placement) [52]. The mean of the three observers was annotated as the ground truth landmark.

This presented DTM method and the two commercial methods were compared to the ground truth landmarks. The analysis was based on two parameters for the calculation of the performance of landmarking algorithms. The first one is a normalised Euclidean distance-error and the second one we call accuracy.

1. Normalised Euclidean distance-error

   The distance-error is the Euclidean distance d between the estimated position coordinates $(\tilde{x}, \tilde{y}, \tilde{z})$, and the actual position (ground-truth) coordinates (x, y, z), of the landmark. For landmarking analysis, the distance-error is divided by the inter-ocular distance (IOD) and multiplied by 100 to correct for scale difference in faces. This results in the normalised distance-error (NDE) in percentage of the IOD.

$$NDE_i^k = 100\frac{d_i^k}{IOD} \tag{3.2}$$

$$d_i^k = \sqrt{(x_i^k - \tilde{x}_i^k)^2 + (y_i^k - \tilde{y}_i^k)^2 + (z_i^k - \tilde{z}_i^k)^2} \tag{3.3}$$

   Here, the superscript k indicates the index of the of the landmarks and the subscript i the index of the patient/subject.

2. Accuracy

   Accuracy is the percentage of estimated landmarks that are detected. A general agreement in the literature is that an estimated landmark is detected if the landmark is within 10% of the IOD of the actual position (ground truth) of the landmark ($NDE_i^k < 10$) [51]. With this criterion we can calculate the accuracy.

   Per-landmark:

$$P(k) = 100\frac{\sum_{i=1}^{I}[i : NDE_i^k < Th]}{I} \tag{3.4}$$

   Here, $[i : NDE_i^k < Th]$ is the indicator function assuming 1, if the deviation is smaller than a threshold (Th), otherwise its value is 0, I denotes the number of subjects.

   Per-subject:

$$P(i) = 100\frac{\sum_{k=1}^{K}[i : NDE_i^k < Th]}{K} \tag{3.5}$$

   Here, K denotes the number of landmarks per subject.

   And the overall performance:

$$P = 100\frac{\sum_{k=1}^{K}\sum_{i=1}^{I}[i : NDE_i^k < Th]}{K \times I} \tag{3.6}$$

The results of this method were compared to two commercially available systems, 3dMD-vultus and Vectra.

Statistical analysis

Statistical data analysis was performed with software package MATLAB version 9.1.0 (MathWorks Inc., Natick, MA, United States).

Intraclass correlation coefficients (ICC) were calculated for the X, Y and Z coordinates of the manually annotated landmarks [53]. This was used as a check for the reliability of the coherence of landmark annotation between the observers. Also, the mean and standard deviation of Euclidian distances between the manually annotated landmarks from the observers, to the average of the observers (ground truth) were calculated, for each landmark. This data was compared to interobserver variabilities known in literature.

The Anderson-Darling test was used to check if the normalised distance error data was normality distributed [54]. The NDE data was divided in four groups and tested individually. The groups are NDE of all landmarks annotated by: (i). the 3dMDVultus system on the 3dMD stereophotographs, (ii). the DTM system on the 3dMD stereophotographs, (iii). the Vectra system on the VectraXT stereophotographs, (iv). the DTM system on the VectraXT stereophotographs. If the null hypothesis of the Anderson-Darling test was not rejected the data was assumed normally distributed. In this case, a paired t-tests should be used to test if there are significant differences between the commercial photo systems landmarking algorithms and the DTM algorithm [55]. If the hypothesis was rejected, the Wilcoxon signed rank test was used for not normally distributed data [56]. The statistical tests are right tailed and performed with a significance level of 5% to reject the hypothesis only in case if the DTM method was significantly better. For cases in which the DTM was not significantly better, a left tail test was performed to check if the 3dMDvultus or Vectra system was performing significantly better. The tests were performed per subject and per landmark. Furthermore, an overall test including all landmarks of all subjects was performed.

The accuracy variable is categorical (true or false) and therefore a the McNemar's test was used to prove statistical differences between landmarking methods. The McNemar's test is a non-parametric test for paired nominal data [57]. The statistical tests are right tailed and performed with a significance level of 5% to reject the hypothesis only in case if the DTM method was significantly better. For cases in which the DTM was not significantly better, a left tail test was performed to check if the 3dMDvultus or Vectra system was performing significantly better. The tests were performed per subject and per landmark. Furthermore, an overall test including all landmarks of all subjects was performed.

3.2 Results

The intraclass correlation coefficients for the X, Y and Z coordinates of all 560 (14 landmarks, 20 subjects and one stereophotograph on each system) landmarks are displayed in table 3.2. Table 3.3 lists the mean and standard deviation of the Euclidian distances between the manually annotated landmarks from the observers, to the ground truth, for each landmark. The average Euclidian distance between all the manually placed landmarks by the three observers (14 landmarks on 40 stereophotographs by 3 observers, total = 1680) and the ground truth was 1.16 mm (SD = 1.88 mm).

Table 3.2: Intraclass correlation coefficients for the X,Y and Z coordinates of all landmarks, and the 95% confidence interval (CI).

ICC (95% CI)	X	Y	Z
Between the 3 observers	0.998 (0.997-0.998)	0.997 (0.997-0.997)	0.999 (0.999-1.000)

Table 3.3: Mean euclidian distances and standard deviations between the manually annotated landmarks of all observers to the ground truth, displayed for each landmark

Landmarks	Δ observer 1,2 and 3 and mean observers (mm) (st.dev. (mm))
1. ex_r	0.70 (0.41)
2. en_r	1.61 (0.45)
3. en_l	1.33 (0.69)
4. ex_l	0.76 (0.50)
5. n	0.70 (0.56)
6. prn	0.48 (0.18)
7. sn	0.76 (0.37)
8. al_r	1.11 (2.27)
9. al_l	1.71 (3.77)
10. ch_r	1.94 (4.14)
11. ch_l	1.31 (2.42)
12. sl	0.75 (0.39)
13. po	2.08 (1.07)
14. g	0.88 (0.32)
Overall	1.16 (1.88)

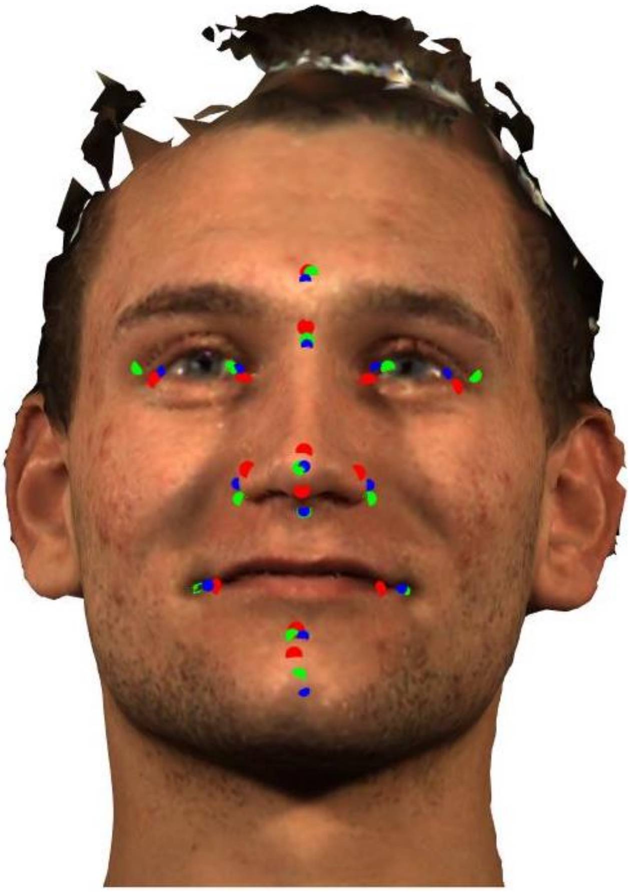
In figure 3.13a the result of a 3D stereophotograph with the annotated landmarks by two automatic systems and the mean of manual annotation by three observers (ground truth). A magnified picture of the nasion landmark is displayed in figure 3.13b where the ground truth is divided in the three landmarks annotated by the independent observers.

Anderson-Darling tests rejected all hypothesis of the NDE data. Therefore, the data was not assumed normally distributed and the Wilcoxon signed rank test was used for statistical testing of the NDE data.

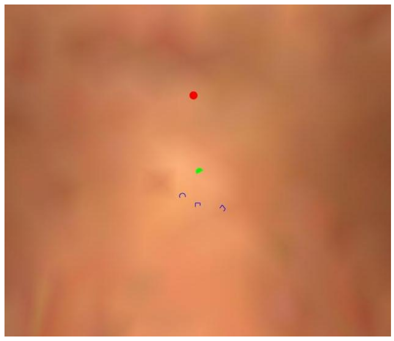
The stereophotographs acquired by the 3dMD system were annotated by the observers, the automatic 3dMDvultus program and the DTM method. The mean normalised distance-error (NDE) in percentage of the IOD by the 3dMDvultus programme over all landmarks was 7.2%, for the DTM method this was 4.7%. The mean IOD was 62.5 mm. The accuracy by the 3dMDvultus programme over all landmarks was 79%, for the DTM method this was 94%. There was a significant difference in the NDE and accuracy of the DTM method and the 3dMDvultus method ($p < 0.05$).

The stereophotographs acquired by the VectraXT system were annotated by the observers, the automatic Vectra program and the DTM method. The Vectra program failed to detect the four eye landmarks for one subject. The NDE mean was calculated without a value for these missed landmarks. The accuracy was calculated with these four landmarks annotated as not detected landmarks. The mean normalised distance-error (NDE) in percentage of the IOD by the Vectra programme over all landmarks was 5.5%, for the DTM method this was 4.0%. The mean IOD was 62.5 mm. The accuracy by the Vectra programme over all landmarks was 86%, for the DTM method this was 97%. There was a significant difference in the NDE and accuracy of the DTM method and the Vectra method ($p < 0.05$).

The NDE values of the 3dMD and DTM method and the Vectra and DTM method per subject are displayed in table 3.4 and per landmark in table 3.6. To evaluate the difference, the p-value of the Wilcoxon signed rank test is shown in table res5 and res7 as well. NDE values are lower with the DTM method compared to the 3dMD software for all subjects and



(a) Results on a 3D stereophotograph



(b) Magnified picture of the nasion landmark

Figure 3.13: Results of automatic annotations and manual annotations on 3D stereophotograph with a magnified picture of the nasion landmark. Red landmarks = Automatically detected landmarks by 3dMDvultus or Vectra system, Green landmarks = Automatically detected landmarks by the DTM method, Blue landmarks = mean of the 3 observers. b. Magnified picture of the nasion landmark with the manually annotated landmarks of the observers shown separately (blue; circle, diamond and square).

landmarks, except subject 5 and 14 and landmark 1, 2 and 4. For the DTM method compared to the Vectra software this lower NDE was seen in all subjects and landmarks, except subject 11 and landmark 1.

The accuracy values of the 3dMD and DTM method and the Vectra and DTM method per subject are displayed in table res6 and per landmark in table res8. To evaluate the difference, the p-value of the McNemar's test is shown in table 3.5 and 3.7 as well. Accuracy values are higher or similar with the DTM method compared to the 3dMD software for all subjects and landmarks, except subject 5 and 14 and landmark 1, 2, 3 and 4. For the DTM method compared to the Vectra software the higher or similar accuracy was seen for each subject and each landmark.

Table 3.4: Normalised distance-error (NDE) results for each patient.

Subjects	3dMD photo acquisition			VectraXT photo acquisition		
	3dMD NDE (%)	DTM NDE (%)	p-value	Vectra NDE (%)	DTM NDE (%)	p-value
1	10.8	5.6	0.029*	5.2	4.8	0.446
2	7.3	3.7	0.068	6.3	4.1	0.047*
3	4.6	3.5	0.251	4.4	3.4	0.122
4	9.2	7.5	0.148	6.2	4.4	0.658
5	4.9	7.8	0.665	4.2	3.8	0.294
6	10.7	2.8	<0.001*	5.0	3.2	0.040*
7	5.5	4.3	0.134	9.9	7.5	0.153
8	10.9	3.4	0.077	5.7	3.9	0.153
9	10.9	4.2	0.003*	6.8	4.1	0.227
10	5.9	4.1	0.045*	4.4	4.0	0.393
11	9.3	7.5	0.357	5.4	7.0	0.830
12	5.1	3.7	0.108	4.5†	3.5	0.150
13	6.5	3.4	0.021*	4.3	2.8	0.108
14	5.8	9.4	0.687	6.2	4.1	0.040*
15	8.0	4.4	0.025*	5.7	3.4	0.011*
16	8.0	4.7	0.039*	8.1	4.2	0.009*
17	6.6	3.6	0.018*	4.3	3.3	0.073
18	4.1	2.8	0.121	4.6	2.7	0.055
19	7.8	4.0	0.018*	4.1	2.7	0.009*
20	5.9	3.5	0.029*	4.1	2.6	0.153
Overall	7.2	4.7	<0.001*	5.5	4.0	<0.001*

*: DTM significantly lower at $p < 0.05$

†: Value of NDE was calculated without a value for the missed four eye landmarks.

Table 3.5: Accuracy results for each patient.

Subjects	3dMD photo acquisition			VectraXT photo acquisition		
	3dMD Acc (%)	DTM Acc (%)	p-value	Vectra Acc (%)	DTM Acc (%)	p-value
1	64	86	0.145	77	92	0.188
2	76	100	0.125	92	100	0.250
3	76	100	0.125	92	100	0.250
4	71	86	0.125	77	100	0.063
5	100	79	0.938	100	100	1.000
6	64	100	0.016*	92	100	0.250
7	86	93	0.313	62	69	0.250
8	86	100	0.125	77	100	0.063
9	57	93	0.035*	92	92	0.500
10	79	100	0.063	85	100	0.125
11	57	79	0.109	85	92	0.250
12	93	100	0.250	69	100	0.031*
13	93	100	0.250	92	100	0.250
14	86	71	0.773	77	100	0.063
15	79	100	0.063	92	100	0.250
16	71	100	0.031*	69	92	0.063
17	71	93	0.109	100	100	1.000
18	93	100	0.250	92	100	0.250
19	71	100	0.031*	100	100	1.000
20	93	100	0.250	92	100	0.250
Overall	79	94	<0.001*	86	97	<0.001*

*: DTM significantly higher at p <0.05

Table 3.6: Normalised distance-error (NDE) results for each landmark.

Landmarks	3dMD photo acquisition			VectraXT photo acquisition		
	3dMD NDE (%)	DTM NDE (%)	p-value	Vectra NDE (%)	DTM NDE (%)	p-value
1. ex_r	3.1	6.2	0.006**	4.2†	6.4	0.006**
2. en_r	3.6	4.7	0.865	3.5†	3.0	0.134
3. en_l	5.7	5.1	0.200	5.2†	3.9	0.034*
4. ex_l	4.2	6.5	0.894	5.3†	4.9	0.619
5. n	5.6	3.8	0.035*	5.3	3.5	0.041*
6. prn	4.7	3.2	0.269	3.3	1.8	0.004*
7. sn	6.3	2.9	0.005*	3.1	2.1	0.038*
8. al_r	10.9	5.4	0.003*	5.5	5.2	0.478
9. al_l	17.6	5.7	<0.001*	8.4	3.8	<0.001*
10. ch_r	8.9	5.2	<0.001*	6.6	3.7	0.023*
11. ch_l	9.0	2.8	<0.001*	5.0	3.3	0.056
12. sl	5.8	4.5	0.113	-	-	-
13. po	11.0	5.1	<0.001*	6.0	5.4	0.052
14. g	4.5	4.4	0.522	9.7	4.6	<0.001*
Overall	7.2	4.7	<0.001*	5.5	4.0	<0.001*

*: DTM significantly lower at $p < 0.05$
**: DTM significantly higher at $p < 0.05$
†: Value of NDE was calculated without a value for the missed eye landmark.

Table 3.7: Accuracy results for each landmark.

Landmarks	3dMD photo acquisition			VectraXT photo acquisition		
	3dMD Acc (%)	DTM Acc (%)	p-value	Vectra Acc (%)	DTM Acc (%)	p-value
1. ex_r	100	85	0.938	85	90	0.313
2. en_r	100	95	0.750	95	100	0.250
3. en_l	100	95	0.750	85	100	0.063
4. ex_l	95	90	0.688	90	95	0.313
5. n	90	95	0.313	90	100	0.125
6. prn	95	100	0.250	100	100	1.000
7. sn	85	95	0.188	100	100	1.000
8. al_r	50	95	0.003*	95	95	1.000
9. al_l	30	90	<0.001*	85	95	0.125
10. ch_r	65	95	0.008*	75	95	0.031*
11. ch_l	70	100	0.008*	85	95	0.125
12. sl	90	90	0.500	-	-	-
13. po	45	95	<0.001*	85	95	0.125
14. g	95	95	0.5	45	100	<0.001*
Overall	79	94	<0.001*	86	97	<0.001*

*: DTM significantly lower at $p < 0.05$

3.3 Discussion

This new DTM method was developed for the automatic detection of landmarks on 3D stereophotographs for training of a prediction model for difficult intubation. The exact features needed for an accurate model are not yet known and therefore the method was chosen in such way that it is highly adaptive. It is possible to add and remove as many landmarks as necessary on the facial template. It is also possible to use the same technique with a different template for different facial postures such as, full mouth opening and neck extension. These extra postures are likely to add significant data for the prediction of difficult intubation. These features are not known in commercial or open-source available landmarking algorithms and therefore, was developed ourselves. The aim of this study was the development and verification of a new landmarking algorithm with these specific requirements. A landmarking system based on deformable template matching was developed and used for verification based on the ground truth (gold standard).

For the ground truth three observers manually annotated the landmarks for both stereophotographs (3dMD and VectraXT) of the 20 subjects. Plooij at al. described a lower mean difference with a comparable standard deviation with the ground truth, and a comparable inter-observer correlation coefficient [52]. Compared to Toma et al. mean distances of our alare, cheilion and glabella landmarks were higher (up to 1mm) and exocantia, subnasale and pronasale of our observers were closer to the ground truth [58]. In the study of Hol et al. carried out by the same department as our study showed comparable inter-observer correlation coefficients and comparable or higher mean distances [59]. Therefore, we conclude that our observer performance according to reliability is comparable to other studies and thus, that our ground truth method is valid.

The overall NDE and accuracy show significant better results between the 3dMDvultus and DTM method, as well as between the Vectra and DTM method. On all patients, similar or significantly better results were seen for each patient. This is also the result for each landmark with the exception of the right exocantion landmark. These results suggest that the DTM method is better for the automatic annotation of facial landmarks on 3D stereophotographs compared to the 3dMDvultus program and Vectra program. Therefore, it can be concluded that the developed algorithm is valid for clinical use. Higher NDE and lower accuracy values were mainly seen in the eye landmarks, especially in the stereophotographs acquired with the 3dMD system, and in one landmark the results were significantly worse. The result of the worse performance of the DTM method was likely due to the use of the template matching and the acquisition method of the 3dMD stereophotographs. Due to the acquisition method and the reflection of the humid eyes the surface mesh is not accurate and relatively course in this area. The presented DTM method is based on the shape of the surface and therefore sensitive to an inaccurate mesh. The 3dMDvultus and Vectra programs make use of the texture map in which the eye landmarks are accurately locatable. The results show that the DTM method could highly benefit of texture based refinement. The initial located landmark could be used as input for a landmark specific texture based refinement algorithm.

A recently published study of de Jong et al. uses a texture map projection of the complete 3D stereophotographs and landmarking based on texture Gabor wavelets [60]. This study showed mean distances of 1 to 2 mm for landmarks easily recognisable on a texture map and distances up to 6 mm for less evident landmarks. Compared to the DTM landmarking method they score higher for the eye and mouth landmarks were our DTM method has better results on nose, chin and eye brow landmarks. This shows that using a combination of a shape based method and a texture based method to could improve automatic landmarking on 3D facial photographs.

The algorithm also located three manually placed green markers. In this study, no quantitative verification of the automatic detection of these markers was performed. Visual evaluation of the automatic detection showed a 100% accuracy of the automatic detected landmarks inside the green markers. Based on these results a quantitative verification at this point of the study was considered of minor importance. In the future, this verification can take place in the same manner as the verification of the DTM method.

A method for the automatic landmark annotation of 3D stereophotographs of the face was presented. With the presented DTM method, it is possible to automatically identify different facial landmarks with acceptable accuracy. The verification was performed to compare this method with two commercially available software systems. It can be concluded that this method is better than those two systems. Furthermore, the method was developed for the analysis of 3D stereophotographs where the located landmarks are used as an input for the training of a prediction model. The accuracy needed for the training of such a model is not yet known and also no validation for this purpose was performed. Texture based refinement of the method, mainly for the eye landmarks, could improve the method significantly. Clinically, the method can be used for the evaluation of 3D stereophotographs based on located landmarks but also complete facial regions could be annotated for analysis.

3.4 Extra clinical applications

The automatic landmarking algorithm which was designed for this study has many other applications besides the landmark localisation for the prediction model. All other studies where 3D facial stereophotograph landmarks are used could benefit from this automatic method. Manual landmarking of 3D stereophotographs is labour-intensive and a lot of research is not performed because there is a lack of time and/or money for landmarking. This method can have a big effect for these studies and can boost a lot of studies based on facial landmarking. These studies include many facial studies about: normal and pathological face growth studies, normal and pathological symmetry studies, effects of various treatments and the planning of treatments to approach normal facial expression.

Not only landmarks, which are vertices on the template mesh, but also other structures on the template could be completely transferred to a 3D stereophotograph. The used template in this study also had some facial esthetic regions annotated. Therefore, stereophotographs can be automatically divided in those regions, after which research based on these regions can be performed. Validation of these regions was performed based on landmarks on the edges of the regions. An article about this method was already submitted [59]. Research about facial regions can include facial growth studies, symmetry studies and follow up studies of all kind of patients.

A specific research field is the evaluation of different facial cancer treatments in children. With the DTM algorithm 3D stereophotographs of children which had different types of treatment (internal radiation or external radiation) were divided in the facial esthetic regions. Area, volume and curvature analysis of these treatments were used to evaluate the treatments. This information will be used to make a prediction for the impaired facial growth for children with cancer. In the future, the information could be used to make a choice about which treatment gives the aesthetically best results. The bases for this research was already done and the DTM algorithm was used to show the possibilities and the first results on patients' 3D stereophotographs. The data of this small pilot was used for a grant application from the KIKA child cancer foundation in the AMC hospital in Amsterdam. The grant was won in April 2017 and the research with facial regions and the use of the DTM method will be continued.

Besides various applications based on facial studies the landmarking method is also applicable of other mesh structures. It is possible to make template meshes of other structures than a face such as bony structures segmented from CT images. With this algorithm, it is possible to automatically place markers or define surfaces on these bones. This could be useful for various principal component analysis or other kind of studies. Overall, this algorithm is useful for the automation of many facial studies and with minor changes a wide spectrum of new study possibilities are within reach. This algorithm and future derivatives of the algorithm will have a significant impact on research and therefore on healthcare.

Chapter 4

Virtual laryngoscopy pilot study

Patients planned for general anaesthesia are evaluated pre-operative for the prediction of difficult intubation. Some difficult intubation predictors are: previous surgery in the head and neck region, space occupying lesions, large deformations in the region and radiotherapy. The prediction is mostly based on assessment of the head and neck region from outside, or history of treatments. This information lacks knowledge about the airway and surrounding tissue itself and therefore an accurate prediction of the difficulty of intubation is challenging. This is information of the middle column as described by the model of Greenland et al. in chapter 2 [5]. Patients undergoing head and neck surgery were frequently reported (39%) in complication cases in the UK. Therefore, these cases require careful airway assessment and planning [2]. Due to pathology or previous surgery in this region a large group of these patients have some kind of medical imaging of this area. CT, MRI or CBCT data could be used to make a 3D virtual laryngoscopy of the airway and therefore has the potential to be of added value for the prediction of a difficult airway. Ahmad et al. described the potential benefit of virtual laryngoscopy but did not support this statement with any kind of research [10]. Cuendet et al. did research about the correlation between virtual and conventional endoscopy. They described a good correlation and a benefit for virtual endoscopy because it is non-invasive, examination of the images is relatively easy, it is cheaper and it is useful for training [11]. Besides these benefits of virtual endoscopy, the 3D image data can also be used to make measurements and gather quantitative data about airway obstructions. In this study, we want to investigate the potential added value of virtual endoscopy and measurements in 3D image data, for the prediction of difficult intubation and help with planning of airway management for suspected patients.

The research question for this path was: Is a virtual laryngoscopy useful for airway planning in patients who are suspected for a difficult airway, due to anatomical changes in or near the airway?

4.1 Methods and Materials

4.1.1 Study cohort

At the beginning of this study the technique of virtual laryngoscopy was presented to a group of anaesthesiologists. During the study these anaesthesiologists could request for a virtual laryngoscopy if there was suspicion of a difficult airway and medical images were available. The OsiriX virtual endoscopy algorithms only input option is CT and therefore in this study we excluded patients with only MR images.

4.1.2 Virtual laryngoscopy

The virtual laryngoscopies were made with advanced DICOM viewer OsiriX (Pixmeo SARL, Bernex, Switzerland). This program has a build in function for a virtual endoscopy. The function was developed for the use of a virtual colonoscopy from CT images but is perfectly usable for other air cavities. The program automatically calculates a path through the air cavity from a manually defined start point to a manually defined endpoint. In our cases the transition from nasopharynx to oropharynx was the starting point and the endpoint was located a few centimetres caudal of the vocal cords. After the calculation of the central luminal path between these two points about 20 to 30 points on the path were automatically selected. For these points, manual selection of the camera orientation was required. During rendering of the virtual endoscopy, the camera orientation of all other points on the path were interpolated. The program has an option to render the endoscopy with or without three 2D orthogonal slices at the position of the camera and the direction of the cameras view. In figure 4.1 an example of a virtual endoscopy with the 2D slices is presented. Together with 3D and 2D visualisation of the segmentation of the airway, the virtual endoscopy was presented to an anaesthesiologist.

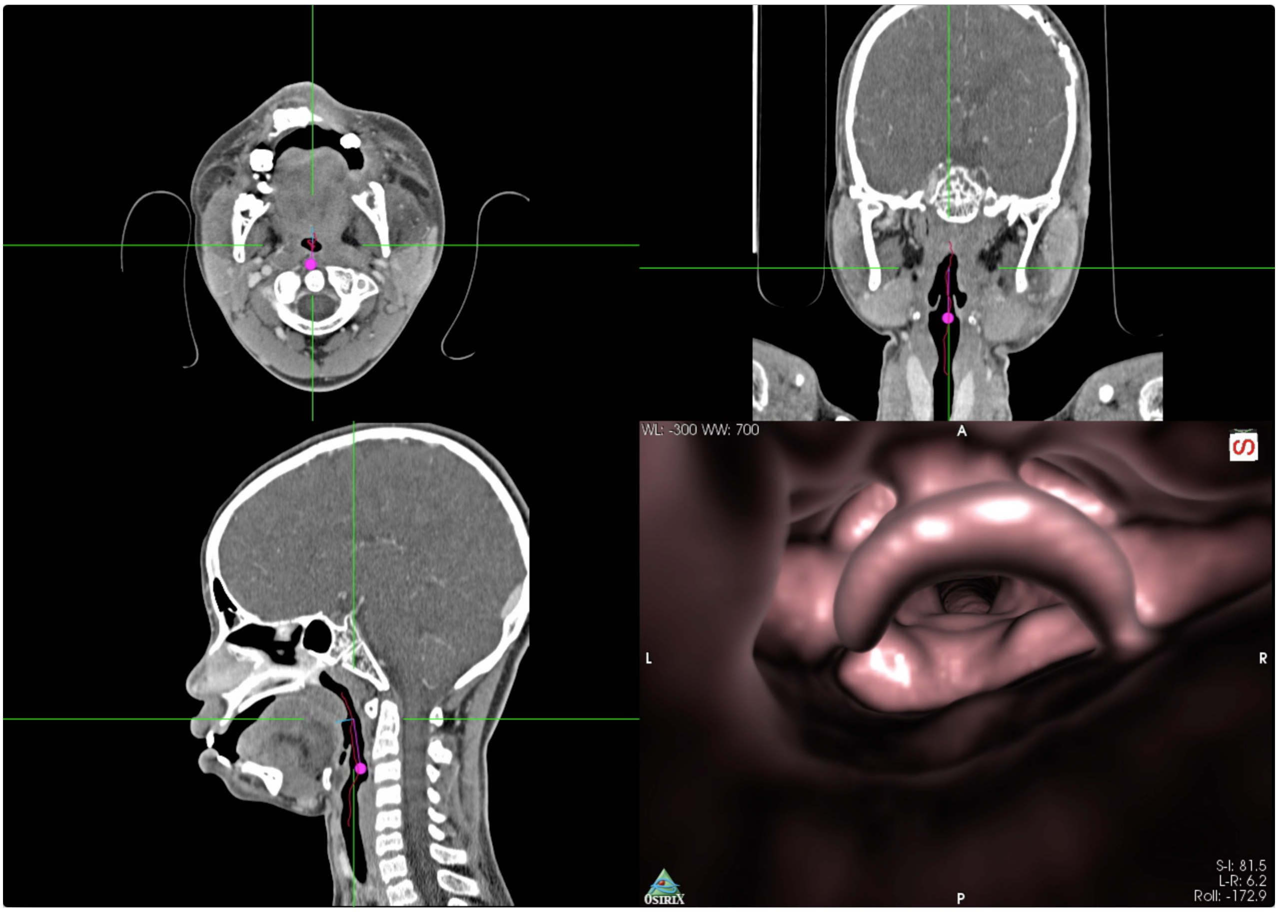


Figure 4.1: Virtual laryngoscopy with 3D orthogonal slices at the camera's position

4.1.3 3D volume and measurements

The program for the visualisation and measurement of the 3D volume was partially developed by the author and adjusted during this study. The algorithm was completely developed in programming software package MATLAB version 9.1.0 (MathWorks Inc., Natick, MA, United States). The program segments the airway with a region growing algorithm designed by Daniel Kellner [61]. After loading of the DICOM dataset a manual crop of the dataset was used to remove all unnecessary data. One point in the airway was manually selected for the region growing. A threshold level of 100 Hounsfield units (HU) and a 26-neighbourhood region growing was used. After segmentation, a 3D Gaussian smoothing operation with a size of 3 was used. An isosurface operation creates a mesh which was used for the visualisation of the airway (Figure 4.2). 2D measurements on the CT data include diameters of the airway at the following levels: cranial of the epiglottis, cranial of the vocal cords, vocal cords and caudal of the vocal cords. If a pathological narrowing of the airway was visible also the diameter and the position of this narrowing was included in the measurements. 2D screenshots and 3D visualisations of this mesh with measurements were used for the presentation of the airway to the anaesthesiologist, together with the virtual laryngoscopy.

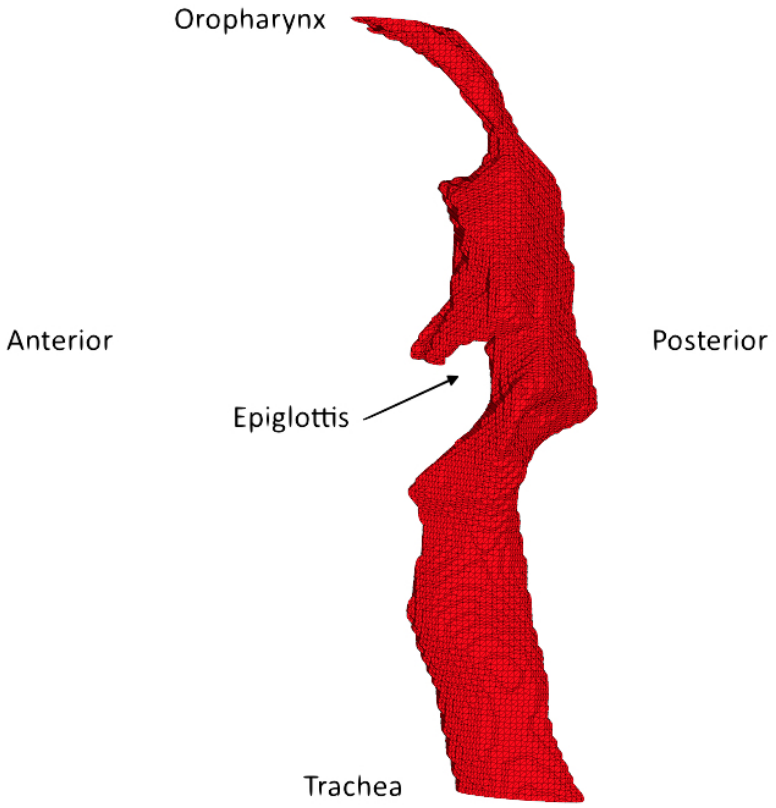


Figure 4.2: 3D visualisation of the segmentation of the airway

4.1.4 Evaluation of the images

The results of the virtual endoscopy, 3D visualisation and the measurements were presented and evaluated with the anaesthesiologists that requested the images, one fixed anaesthesiologist and the head of the anaesthesiologist department. Advantages and disadvantages were discussed.

4.2 Results

During the time of the study, four requests for virtual scopies were handed in and all have been processed. Reasons for request were suspicion of a difficult airway due to previous surgery, suspicion of narrowing due to space occupying lesions in surrounding tissue and extreme scoliosis of the cervical and thoracic vertebrae. One virtual endoscopy was made of an airway without any abnormalities as a reference for a normal 3D virtual endoscopy of the airway.

Basic structures in the virtual endoscopy were relatively easy to find for the untrained anaesthesiologist and also general understanding of the orientation was experience positive. On the other hand, evaluation of only the three orthogonal slices of a normal CT was experienced as difficult and failed in some cases.

Evaluation resulted in a good understanding of the seriousness of a stenosis if one was present. This was reported more easy on the 3D endoscopy compared to the 2D CT slices. Differences in symmetry were easily detected and also the position of a stenosis in the airway was easy to describe.

Patients with large deformations or pathologies in the airway were sometimes harder to evaluate if there was a lack of recognisable landmarks for orientation.

Reported disadvantages of the virtual laryngoscopies included: inability to change the brightness, inability to change the camera orientation after rendering of the movie, real time measurements during evaluation of the endoscopy, no visualisation of the camera position and view in a 3D overview of the airway and the inability to use other imaging modalities.

4.3 Discussion

The aim of this study was to get a basic understanding of the predictive value of a virtual laryngoscopy in patients with a suspicion of difficult intubation because of various reasons in the laryngeal area. Commercially available software with basic features was used for this study. The algorithm was not developed for this specific goal but it was useful to get a general understanding about the predictive value of a virtual endoscopy. One healthy example and four patients were used in this study to evaluate the usefulness of virtual laryngoscopy. Analysis on these five virtual endoscopies resulted in advantages of this technique. Evaluation resulted in a set of extra features that need to be implemented in a dedicated software package for this specific purpose.

The biggest advantage of a virtual laryngoscopy for anaesthesiologist is the 3D view and the orientation of the camera. This orientation is the same compared to a direct laryngoscopy when performed during intubation by an anaesthesiologist. The lack of experience with the evaluation of 2D medical imaging makes it difficult for an anaesthesiologist to estimate the seriousness of a stenosis and the possible effect during intubation. In 3D it is relatively easy to form an opinion about the amount of space or lack of space in the airway. Future implementation and standard use of this technique will result in even more understanding of an easy or difficult to intubate airway based on the 3D virtual laryngoscopy.

Based on this study a number of features need to be develop in a dedicated software package before the technique can be used optimally. The software should be compatible with the use of CT, MRI and CBCT images as input. The change of brightness of virtual endoscopy is necessary to evaluate darker areas. The virtual laryngoscopy should not be a simple movie but an interactive program. In this program, it must be possible to move the camera over a central luminal line and manually adjust the cameras orientation in real time. If no adjustment of the camera is done, the basic orientation should be in the direction of the central luminal line. An extra window with a 3D visualisation of the airway, position and orientation of the camera should be added. This improves the understanding of position in the body and orientation of the camera. Finally, it must be possible to perform some measurements in real-time by clicking on structures visible in the virtual endoscopy.

During direct laryngoscopy for intubation of a patient, the patient receives muscle relaxant drugs. These drugs are necessary for surgery and can also help during intubation. The disadvantage of the use of muscle relaxants is the change in anatomy in the head and neck region compared to the anatomy in an awake patient. The images used for the 3D virtual laryngoscopy were acquired in an awake patient and therefore could differ form the situation during intubation. Better understanding of the effect of muscle relaxants and the effect on the anatomy for intubation is needed. At the Radboud UMC a study will be started to fully understand this effect. If the effect is better understood it may be possible to correct for this effect in the virtual images.

Virtual laryngoscopy is a useful tool for the evaluation of medical images without extensive knowledge about the interpretation of conventional medical images in 2D views. Therefore, it is an excellent way to evaluate the airway for anaesthesiologist who normally lack large experience in the interpretation of 2D medical images.

Chapter 5

Conclusions and future prospects

5.1 Conclusions

The general aim of this study was to improve the prediction of difficulty during intubation of patients during general anaesthesia. Difficult airway prediction and airway management is studied for many years and a lot of methods are developed to predict a difficult airway. Unfortunately, all prediction methods are complicated or little sensitive and little specific [8]. An easy and accurate prediction of a difficult airway can prepare anaesthesiologists to prevent serious complications and stressful situations during general anaesthesia.

During this one year graduation project, two different methods to improve the difficult airway prediction were studied. Both involving 3D imaging technologies. The first method is a new way for the initial prediction of a difficult airway with a computerised model based on 3D facial stereophotographs. Nowadays, this prediction is done at a pre-operative consult by the subjective eye of an anaesthesiologist with the help of some standard parameters and tests. We wanted to objectify and standardise the initial prediction of a difficult airway with a model made with machine learning. This method was divided in two studies: (i) Literature research to evaluate the documentation needed for the development of this model and to facilitate airway management evaluation; (ii) An analysis tool for automatic detection of landmarks on 3D stereophotographs to use as input features to train the prediction model. The second method and third study (iii) aims on patients with a suspicion of a difficult airway with already available 3D imaging of the head and neck region. Nowadays, medical imaging is hardly used by the anaesthesiology department while there is a possibility that they could benefit from this information. The use of virtual endoscopy algorithms transform the imaging data form 2D to 3D which makes it a lot easier to evaluate by an inexperienced observer. This study was a pilot to test the added value of a virtual laryngoscopy for the anaesthesiology department.

(i). The literature study resulted in an answer for the question: Which changes need to be made in the documentation of airway management at the Radboud UMC, in order to gather data for a baseline study and monitoring of the performance of airway management. Pre-operative documented parameters for the prediction of a difficult airway were not changed because all parameters with the highest possible sensitivity and specificity were already standard for documentation. Already documented pre-operative parameters are: Mallampati score, TMD, mouth opening, neck extension and neck circumference. The only addition to the pre-operative documentation included a final verdict of the anaesthesiologist about the predicted difficulty, yes or no. This is necessary for a baseline study to calculate the performance of the current prediction method and for monitoring possible changes in the method in the future. Intra-operative changes were made in the documentation of the used airway management methods. Already documented parameters about intubation were: Cormack-Lehane score. Added mandatory fields for documentation are: number of attempts, number of operators, use of extra materials or techniques and the change of primary airway plan. The additional parameters will be used for the discrimination between easy or difficult intubation. The definition of easy and difficult intubation varies between studies and mainly is a specific combination and scores on these parameters. With the addition of these parameter we can divide our patients in the same way as most other studies in literature do.

(ii). The development of an airway difficulty prediction model needed a tool for the automatic detection of facial features. A study was started for the development of such a tool and the verification of the tool to answer the research question: what is the performance of a new developed deformable template matching landmark detection algorithm for use with 3D stereophotographs? The verification was done by comparing the new method with two commercially available systems, each with 20 subjects. It can be concluded that the presented DTM method overall performance was significantly better ($p < 0.001$) than the two commercial systems. This conclusion was made based on two parameters: the normalised distance-error and accuracy.

(iii). The research question for the last pathway was: Is a virtual laryngoscopy useful for airway planning in patients who are suspected for a difficult airway, due to anatomical changes in or near the airway? The study used an already develop software program for the virtual laryngoscopy with limited options. Four patients were used for the evaluation of the method, three retrospective and one patient was not yet intubated at the end of the study. However solid conclusions about the added value of virtual laryngoscopy could not be made after this small pilot study, the results from anaesthesiologists are promising and there is a lot of enthusiasm for this new prediction modality.

Overall, the three research pathways were started to investigate the benefits of 3D imaging for the prediction of a difficult airway. This is desirable to lower the incidence of airway related complications and the corresponding severe risks for a patient, like: brain damage and death. During the time of the studies some benefits were already seen and more possibilities were investigated. Although, more time and research is needed to finish some paths and draw more solid conclusions, it can be stated that 3D imaging for the prediction of a difficult airway is promising for the improvement of healthcare by the department of anaesthesiology. 3D imaging is already an improvement in other departments of the hospital and it is also promising to have a significant impact on healthcare of the department of anaesthesiology for the prediction of a difficult airway.

5.2 Future prospects

As stated in the first chapter, this research was divided in three parallel pathways to investigate the possible benefit of 3D imaging for the prediction of a difficult airway. Within the time of these studies some conclusions were made and some predictions were done. Further research is necessary to prove or disprove these predictions and to finish the research that is started.

After implementation of the changes made in the pre- and intra-operative parameters for documentation, a start needs to be made for the baseline study. With this study objective performance measurements of the current prediction method can be made. This is useful for comparison with other prediction models like the computerised model that is planned for the near future. The documentation is also useful to continue after implementation of changes in the prediction of a difficult airway, to monitor the results of these changes. Besides a baseline study also a computerised prediction model based on the intra-operative manually evaluated parameters can be made. At the moment, there is no standardised method for the evaluation of the individual parameters and how to reach a final verdict based on the five individual parameters. Machine learning could be used to make a model only on these five parameters.

For the development of a computerised prediction model based on 3D stereophotographs, a 3D facial analysis tool was made. A lot of improvements are possible for this tool, but without these improvements the tool is almost ready for the analysis of stereophotographs for the training of a prediction model. The 3D stereophotographs that will be used as input for the model are made in three different facial postures. The algorithm was only tested on one pose and two more templates are needed for the other two postures. During implementation of the addition of two extra templates the acquisition of a database of 3D stereophotographs should be started. After surgery and with the newly implemented intra-operative parameters, the patient with corresponding 3D stereophotographs, can be divided one of two groups: easy-to-intubate and difficult-to-intubate. After finishing the analysis tool and a database of at least 80 easy-to-intubate and 80 difficult-to-intubate patients, a prediction model can be made with machine learning techniques. A 40-40 group can be used for training and 40-40 for validation. Such models have a threshold which can be chosen to classify an output of the model as easy or difficult to intubate. In a receiver operating characteristic curve the effect of the chosen threshold on the specificity and sensitivity is visualised. Thereby, it is possible to choose the percentage of false positives and corresponding false negatives. In the current chosen airway perdition method, this is only possible by different training of anaesthesiologists and individual experience. Computerised prediction models could improve and objectify the prediction of a difficult airway. Computerised models are also easier to adjust for different performance like higher sensitivity and lower specificity if desired. These models could have a significant impact on healthcare due to the easiness and speed of accurate measurements replacing guessing and individual differences in experience.

There are improvements possible for the DTM method. For almost all sub functions there are already ideas for improvements. The main improvement can be made in calculation time. The complete algorithm calculation time for one stereophotograph was about 10 minutes. In a batch run this was 10 minutes for four facial 3D stereophotographs. The CPD algorithm takes about 95% of this time and therefore the biggest improvement is needed in that algorithm. The calculation time can be reduced by using smaller mesh sizes. Research is needed to investigate how much a facial stereophotograph can be reduced in vertices in order to increase speed without the loss of landmark accuracy. A CPD algorithm that makes use of the graphics processing unit (GPU) could significantly increase speed also. An other speed improvement could be a change in programming language. C++ programming could highly increase algorithm speed but is only effective if the development of the complete algorithm is at its end stage. This is due to de easiness of development and more general knowledge of employees about the MATLAB environment compared to C++ programming.

The prediction of difficult intubation can also benefit of ultrasound, CT, CBCT and MRI distances measurements [10,62–64]. Future research could include these measurements in a more advanced model. There is a possibility to use a model based on only 3D stereophotographs and after an outcome of difficult intubation, use extra imaging like ultrasound to add data to the model and increase the predictive value. Research is needed to investigate if other image modalities can indeed improve a prediction model compared to a model based on 3D stereophotographs only.

The continuation of the virtual laryngoscopy study has already started. A start is made with the development of dedicated software for virtual laryngoscopies. This software package will contain improvements as indicated in the discussion of the study. With this new improved software package a larger study needs to be started. A prospective study involving more patients. Evaluations of the virtual laryngoscopies can be documented and during or after sugary checked with the actual findings. Also, monitoring of the laryngoscopy with a video laryngoscope is possible for comparison with the virtual laryngoscopy.

Bibliography

[1] Connor CW, Segal S. *Accurate classification of difficult intubation by computerized facial analysis.* Anesthesia and Analgesia. 2011;**112**(1):84–93.

[2] Cook TM, Woodall N, Frerk C. *Major complications of airway management in the UK: Results of the Fourth National Audit Project of the Royal College of Anaesthetists and the Difficult Airway Society. Part 1: Anaesthesia.* British Journal of Anaesthesia. 2011;**106**(5):617–631.

[3] Nagaro T, Yorozuya T, Sotani M, et al. *Survey of patients whose lungs could not be ventilated and whose trachea could not be intubated in university hospitals in Japan.* Journal of Anesthesia. 2003;**17**(4):232–240.

[4] Eindhoven GB. *14 Luchtwegmanagement.* In: Anesthesiologie, PJ Hennis. Bohn Stafleu van Loghum, Houten; 2007. p. 155–166.

[5] Greenland KB. *Airway assessment based on a three column model of direct laryngoscopy.* Anaesthesia and Intensive Care. 2010;**38**(1):14–19.

[6] Rose DK, Cohen MM. *The incidence of airway problems depends on the definition used.* Canadian Journal of Anaesthesia. 1996;**43**(1):30–34.

[7] Cormack RS, Lehane J. *Difficult tracheal intubation in obstetrics.* Anaesthesia. 1984;**39**(11):1105–1111.

[8] Yentis SM. *Predicting difficult intubation - Whorthwhile exercise or pointless ritual?* Anaesthesia. 2002;**57**:105–109.

[9] Cuendet G, Schoettker P, Yuce A, et al. *Facial Image Analysis for Fully-Automatic Prediction of Difficult Endotracheal Intubation.* IEEE transactions on bio-medical engineering. 2015;**63**(2):328–339.

[10] Osorio F, Perilla M, Doyle DJ, et al. *Cone beam computed tomography: An innovative tool for airway assessment.* Anesthesia and Analgesia. 2008;**106**(6):1803–1807.

[11] Ahmad I, Millhoff B, John M, et al. *Virtual endoscopy—a new assessment tool in difficult airway management.* Journal of Clinical Anesthesia. 2015;**27**(6):508–513.

[12] Benumof JL. *Difficult laryngoscopy: obtaining the best view.* Canadian Journal of Anaesthesia. 1994;**41**(1):361–365.

[13] Matsumoto T, Carvalho WBD. *Tracheal intubation.* Jornal de Pediatria. 2007;**83**:83–90.

[14] Greenland KB. *A proposed model for direct laryngoscopy and tracheal intubation.* Anaesthesia. 2008;**63**(2):156–161.

[15] Caplan R, Benumof J, Berry F, et al. *Practice guidelines for management of the difficult airway.* Anesthesiology. 1993;**78**:597–602.

[16] Mallampati SR. *Clinical sign to predict difficult tracheal intubation (hypothesis).* Canadian Anaesthetists' Society Journal. 1983;**30**(3):316–317.

[17] Mallampati SR, Gatt SP, Gugino LD, et al. *A clinical sign to predict difficult tracheal intubation; a prospective study.* Canadian Anaesthetists' Society Journal. 1985;**32**(4):429–434.

[18] Samsoon GLT, Young JRB. *Difficult tracheal intubation: a retrospective study.* Anaesthesia. 1987;**42**(5):487–490.

[19] Lee A, Fan LTY, Gin T, et al. *A systematic review (meta-analysis) of the accuracy of the mallampati tests to predict the difficult airway.* Anesthesia and Analgesia. 2006;**102**(6):1867–1878.

[20] Shiga T, Wajima Z, Inoue T, et al. *Predicting Difficult Intubation in Apparently Normal Patients.* Anesthesiology. 2005;**103**(2):429–437.

[21] Domi R. *The best prediction test of difficult intubation.* Journal of Anaesthesiology Clinical Pharmacology. 2010;**26**(2):193–196.

[22] Lundstrøm LH, Vester-Andersen M, Møller AM, et al. *Poor prognostic value of the modified Mallampati score: A meta-analysis involving 177 088 patients.* British Journal of Anaesthesia. 2011;**107**(5):659–667.

[23] Available from: http://cursoenarm.net/UPTODATE/contents/mobipreview.htm?19/30/19942; [cited 19-04-2017].

[24] Baker PA, Depuydt A, Thompson JMD. *Thyromental distance measurement - Fingers don't rule.* Anaesthesia. 2009;**64**(8):878–882.

[25] Available from: https://sites.google.com/site/rmhicucredentialing/endotracheal-intubation/airway-assessment/thyromental-distance; [cited 19-04-2017].

[26] Rose DK, Cohen MM. *The airway: problems and predictions in 18,500 patients.* Canadian Journal of Anaesthesia. 1994;**41**(5 Pt 1):372–83.

[27] Juvin P, Lavaut E, Dupont H, et al. *Difficult Tracheal Intubation Is More Common in Obese Than in Lean Patients.* Anesthesia and Analgesia. 2003;**97**:595–600.

[28] El-Ganzouri AR, McCarthy RJ, Tuman KJ, et al. *Preoperative airway assessment: predictive value of a multivariate risk index.* Anesthesia and analgesia. 1996 jun;**82**(6):1197–204.

[29] Khan ZH, Kashfi A, Ebrahimkhani E. *A comparison of the upper lip bite test (a simple new technique) with modified Mallampati classification in predicting difficulty in endotracheal intubation: A prospective blinded study.* Anesthesia and Analgesia. 2003;**96**(2):595–599.

[30] Eberhart LHJ, Arndt C, Cierpka T, et al. *The reliability and validity of the Upper Lip Bite test compared with the Mallampati classification to predict difficult laryngoscopy: An external prospective evaluation.* Anesthesia and Analgesia. 2005;**101**(1):284–289.

[31] Myneni N, O'Leary AM, Sandison M, et al. *Evaluation of the upper lip bite test in predicting difficult laryngoscopy.* Journal of Clinical Anesthesia. 2010;**22**(3):174–178.

[32] Hester CE, Dietrich SA, White SW, et al. *A comparison of preoperative airway assessment techniques: The modified Mallampati and the upper lip bite test.* AANA Journal. 2007 jun;**75**(3):177–182.

[33] Woodward LJ, Kam PCA. *Ankylosing spondylitis: Recent developments and anaesthetic implications.* Anaesthesia. 2009;**64**(5):540–548.

[34] Nasa VK, Kamath SS. *Risk factors assessment of the difficult intubation using intubation difficulty scale.* Journal of Clinical and Diagnostic Research. 2014;**8**(7):10–12.

[35] Huh J, Shin HY, Kim SH, et al. *Diagnostic predictor of difficult laryngoscopy: The hyomental distance ratio.* Anesthesia and Analgesia. 2009;**108**(2):544–548.

[36] Gonzalez H, Minville V, Delanoue K, et al. *The importance of increased neck circumference to intubation difficulties in obese patients.* Anesthesia and Analgesia. 2008;**106**(4):1132–1136.

[37] Kim WH, Ahn HJ, Lee CJ, et al. *Neck circumference to thyromental distance ratio: A new predictor of difficult intubation in obese patients.* British Journal of Anaesthesia. 2011;**106**(5):743–748.

[38] Nørskov AK, Wetterslev J, Rosenstock CV, et al. *Effects of using the simplified airway risk index vs usual airway assessment on unanticipated difficult tracheal intubation - a cluster randomized trial with 64,273 participants.* British Journal of Anaesthesia. 2016 may;**116**(5):680–689.

[39] Eberhart LH, Arndt C, Aust HJ, et al. *A simplified risk score to predict difficult intubation: development and prospective evaluation in 3763 patients.* Eur J Anaesthesiol. 2010;**27**(11):935–940.

[40] Shah P, Sundaram V. *Incidence and predictors of difficult mask ventilation and intubation.* Journal of Anaesthesiology Clinical Pharmacology. 2012;**28**(4):451.

[41] Naguib M, Scamman FL, O'Sullivan C, et al. *Predictive performance of three multivariate difficult tracheal intubation models: A double-blind, case-controlled study.* Anesthesia and Analgesia. 2006;**102**(3):818–824.

[42] Wilson ME, Spiegelhalter D, Robertson JA, et al. *Predicting difficult intubation.* British Journal of Anaesthesia. 1988;**61**(2):211–216.

[43] Langeron O, Cuvillon P, Ibanez-Esteve C, et al. *Prediction of difficult tracheal intubation: time for a paradigm change.* Anesthesiology. 2012;**117**(6):1223–33.

[44] Law JA, Broemling N, Cooper RM, et al. *The difficult airway with recommendations for management - Part 1 - Intubation encountered in an unconscious/induced patient.* Canadian Journal of Anesthesia. 2013;**60**(11):1089–1118.

[45] Apfelbaum JL, Hagberg Ca, Caplan Ra, et al. *Practice Guidelines for Management of the Difficult Airway.* Anesthesiology. 2003;**98**(2):1269–1277.

[46] Adnet F, Borron SW, Racine SX, et al. *The intubation difficulty scale (IDS): Proposal and evaluation of a new score characterizing the complexity of endotracheal intubation.* Anesthesiology. 1997;**87**(6):1290–1297.

[47] Myronenko A, Song X. *Point set registration: Coherent point drifts.* IEEE Transactions on Pattern Analysis and Machine Intelligence. 2010;**32**(12):2262–2275.

[48] Jia Y, Shelhamer E, Donahue J, et al. *Caffe: Convolutional Architecture for Fast Feature Embedding.* IEEE transactions on neural networks. 2014 jun;**5**(2):157–66.

[49] Zhang K, Zhang Z, Li Z, et al. *Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks.* IEEE Signal Processing Letters. 2016 oct;**23**(10):1499–1503.

[50] Gower JC. *Generalized procrustes analysis.* Psychometrika. 1975;**40**(1):33–51.

[51] Çeliktutan O, Ulukaya S, Sankur B. *A comparative study of face landmarking techniques.* EURASIP Journal on Image and Video Processing. 2013 dec;**2013**(1):13.

[52] Plooij JM, Swennen GRJ, Rangel FA, et al. *Evaluation of reproducibility and reliability of 3D soft tissue analysis using 3D stereophotogrammetry.* International Journal of Oral and Maxillofacial Surgery. 2009;**38**(3):267–273.

[53] Shrout PE, Fleiss JL. *Intraclass correlations: Uses in assessing rater reliability.* Psychological Bulletin. 1979;**86**(2):420–428.

[54] Anderson TTW, Darling Da. *Asymptotic theory of certain" goodness of fit" criteria based on stochastic processes.* The annals of mathematical statistics. 1952;**23**(2):193–212.

[55] Student. *The probable error of a mean.* Biometrika. 1908;**6**(1):1–25.

[56] Wilcoxon F. *Individual Comparisons by Ranking Methods.* Biometrics Bulletin. 1945;**1**(6):80.

[57] McNemar Q. *Note on the sampling error of the difference between correlated proportions or percentages.* Psychometrika. 1947;**12**(2):153–157.

[58] Toma AM, Zhurov A, Playle R, et al. *Reproducibility of facial soft tissue landmarks on 3D laser-scanned facial images.* Orthodontics and Craniofacial Research. 2009;**12**(1):33–42.

[59] Hol MLF, Meulstee JW, Merks JM, et al. *Analyzing 3D stereophotographs: a new method for the automatic segmentation of the complete human face into facial esthetic units.* submitted to PLOS ONE. 2017;.

[60] De Jong MA, Wollstein A, Ruff C, et al. *An automatic 3D facial landmarking algorithm using 2D gabor wavelets.* IEEE Transactions on Image Processing. 2016;**25**(2):580–588.

[61] Available from: https://nl.mathworks.com/matlabcentral/fileexchange/32532-region-growing--2d-3d-grayscale-?focused=5195969&tab=function; [cited 09-05-2017].

[62] Samra SK, Schork MA, Guinto FC. *A study of radiologic imaging techniques and airway grading to predict a difficult endotracheal intubation.* Journal of Clinical Anesthesia. 1995;**7**(5):373–379.

[63] Naguib M, Malabarey T, AlSatli RA, et al. *Predictive models for difficult laryngoscopy and intubation. A clinical, radiologic and three-dimensional computer imaging study.* Canadian journal of anaesthesia = Journal canadien d'anesthésie. 1999;**46**(8):748–59.

[64] Mao X, Xiaoxi L, Jun W, et al. *Application of a new combined model including radiological indicators.* Chinese Medical Journal. 2014;**127**(23):4043–4048.

Appendix A

Graduation resume

This appendix containts a short summary of other project and activities I have worked on this year, besides the studies explained in this report.

Pilot study analysis of facial 3D stereophotographs of children treated for cancer for the application of a KIKA grant.

For the application of a KIKA grant an analysis method was developed by automatic segmentation of facial regions with the developed DTM method of chapter 3. After automatic segmentation of various 3D stereophotographs, facial regions were individually analyzed by area, volume and curvature. The project contained al analysis and programming for this analysis in MATLAB. The grant was won in April 2017.

Growth study between four groups divided by age (11-13 years and 17-19 years) and gender.

Facial 3D photographs were analysed with the DTM method of chapter 3 to investigate facial growth between the ages and de differences between boys and girls. The group sizes will be enlarged before the final conclusions can be made. Eventually this project will contain 3D stereophotographs of children from 1 to 20 years and an extensive facial growth study will be performed.

Writing of a CMO request for the acquisition of 3D stereophotographs for the development of a difficult airway prediction model for the department of anaesthesiology.

As stated in this report a model will be trained for the prediction of a difficult airway. Therefore, data acquisition of patients planned for surgery needs to be set up. Many logistical and legal steps for this acquisition were taken this year and a CMO request has been submitted for approval.

Development of an analysis method for the evaluation of eyelid correction surgery based on 3D stereophotographs.

An analysis method is developed and programmed in MATLAB for a surgeon of the Spaarne Gasthuis Hospital in Haarlem. The surgeon wanted to evaluate the effects of eyelid correction and forehead lifts on the volume of the eyelid and the location of the eyebrow. Therefore, he already made 3D stereophotographs pre- and post-operative. This year we started with a validation study of the eyelids on 3D stereophotographs to check the reputability of these structures during the day. A method for data acquisition, and evaluation was developed and programmed. The study will continue with the evaluation of the data and the analysis of the patient's 3D stereophotographs.

Development of a validation study for the evaluation of eyelid correction surgery based on high-resolution 3D scanner images.	A data acquisition study is designed for the validation of the repeatability of the eye region on 3D images made with a high-resolution 3D scanner. This validation is needed for a study about the effects of eyelid correction surgery measured by these images. This research is done for a surgeon at the university medical center in Groningen.
Minor work for the 3D lab of the Oral and Maxillofacial surgery department of the Radboud UMC	During the year, all kinds of clinical 3D analysis and planning work was performed on CBCT scans and 3D stereophotographs. This includes: orbital volume measurements, mirroring of facial images for the predicted outcome of treatment and the planning of implants and 3D reconstructions of the head and neck region for pre-operative planning or post-operative evaluation.

Appendix B

Procrustus algorithm

The Procrustes algorithm is a method which can be used to calculate the linear transformation (rotation, scale, translation and reflection components) to align two point clouds [50]. The point clouds have to be equal in size and the individual points should have the same order in both point clouds. After calculation of the scale S, rotation matrix $\mathbf{R}$ (with or without reflection possibility) and the translation vector $\mathbf{t}$, between two point clouds the same parameters can be used on a different point cloud for the linear transformation. The input of the Procrustes algorithm are two point clouds with corresponding points $\mathbf{x}_n$ and $\mathbf{y}_n$ in $3 \times N$ matrices $\mathbf{X}$ and $\mathbf{Y}$ for 3D point clouds . The model for alignment of points $\mathbf{y}_n$ to $\mathbf{x}_n$ is $\mathbf{x}_n = S\mathbf{R}\mathbf{y}_n + \mathbf{t}$. The output of the Procrustes algorithm is: Scale S, rotation matrix $\mathbf{R}$, and translation vector $\mathbf{t}$.

The Procrustes algorithm uses the following steps:

1. Normalization with respect to the centroids of the point clouds:

   - Calculate centroids of datasets:

$$\mathbf{x} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{x}_n \quad \text{(B.1)} \qquad \mathbf{y} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{y}_n \quad \text{(B.2)}$$

   - Normalize for all n:

$$\tilde{\mathbf{x}}_n = \mathbf{x}_n - \bar{\mathbf{x}}_n \quad \text{(B.3)} \qquad \tilde{\mathbf{y}}_n = \mathbf{y}_n - \bar{\mathbf{y}}_n \quad \text{(B.4)}$$

2. Calculate and neutralize the scale S:

   - Calculate the scale:

$$\hat{S} = \frac{\sum_{n=1}^{N}\|\mathbf{x}_n\|}{\sum_{n=1}^{N}\|\mathbf{y}_n\|} \tag{B.5}$$

   - Rescale $\tilde{\mathbf{x}}_n$:

$$\tilde{\mathbf{x}}_n = \frac{1}{S}\tilde{\mathbf{x}}_n \tag{B.6}$$

3. Calculate the rotation matrix $\mathbf{R}$ using Kabsch algorithm:

   - Singular value decomposition (SVD): calculate output $\mathbf{U}$,$\mathbf{S}$,$\mathbf{V}$, such that:

$$\mathbf{U}\mathbf{S}\mathbf{V}^T = \tilde{\mathbf{X}}\tilde{\mathbf{Y}}^T \tag{B.7}$$

   - Calculate rotation matrix $\mathbf{R}$:

$$\hat{\mathbf{R}} = \mathbf{U}\mathbf{W}\mathbf{V}^T \tag{B.8}$$

- Where matrix $\mathbf{W}$, with the determinant of $\mathbf{U}\mathbf{V}^T$ is used to correct for reflection if one is present and not allowed:

$$\mathbf{W} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & det(\mathbf{U}\mathbf{V}^T) \end{bmatrix} \tag{B.9}$$

4. Calculate the translation vector:

$$\hat{\mathbf{t}} = \bar{\mathbf{x}} - \hat{S}\hat{\mathbf{R}}\bar{\mathbf{y}} \tag{B.10}$$

After calculation of the scale, rotation matrix and translation vector the transformation can be applied to other point clouds with the following formula:

$$\mathbf{z}'_n = S\mathbf{R}\mathbf{z}_n + \mathbf{t} \tag{B.11}$$

Here, $\mathbf{z}'_n$ are the transform points of the to be transformed point cloud $\mathbf{z}_n$. The scaling can simply be ignored by leaving the scale S out of formula A.11. Reflection can be allowed by leaving the matrix $\mathbf{W}$ out of equation A.8.