

TECHNICAL MEDICINE

BACHELOR THESIS

---

# Eliminating global head movement during swallowing

to create a more objective, accurate and reproducible assessment of the swallowing function using videofluoroscopy

---

*Authors:*

M.H. GERKEMA

M.H.M KETEL

E.P.J. SMITS

M.D. van der STOEL

*Supervisors:*

Prof. dr. A.J.M. BALM

Dr. ir. F. van der HEIJDEN

Dr. L. van der MOLEN

UNIVERSITY OF TWENTE.



June 24, 2015



# Abstract

**Aim:** The standard diagnosis-tool for dysphagia is videofluoroscopy (VFS). At the moment, the assessment of the videofluoroscopy videos using the Logemann-protocol is subjective and time-consuming. The aim of this research is to make a start with developing a (semi)-automatic method to assess VFS, by eliminating the global head movement of subjects during swallowing. These results will eventually contribute to making the assessment of videofluoroscopy more objective, reproducible, accurate and efficient.

**Method:** We received VFS videos of eight healthy subjects from the Antoni van Leeuwenhoek hospital in Amsterdam. Of these videos, first fragments were selected and subsequently five frames, showing different stages of the swallowing process. The demons algorithm is used to match the frames to each other. Using MATLAB<sup>®</sup>, we placed landmarks on stable anatomical structures, using three different methods: landmarks < 30, landmarks > 30 and landmarks placed on the jaw. After obtaining a set of landmarks in a reference and moving frame, the frames were matched by the mesh-based warping algorithm. Further on, landmarks were also obtained by segmenting the structures, which produced larger amounts of landmarks utilised in the algorithms. The segmented frames were also matched with the mesh-based algorithm. After the transformation the frames were computed in a video.

**Results:** In the frames transformed with the demons algorithm, the vertebrae are not yet entirely stabilized. The demon-transformed frames also show deformation of the anatomical structures and of the swallowing process. Of the frames transformed with the mesh-based warping algorithm using landmarks, the method with landmarks on the jaw show the best results. The head became almost motionless and in most cases deformation of the swallowing function was limited. In the frames transformed with mesh-based warping using segmentation, the movement of the head was almost eliminated. Although, the movement differed intersubjectively, the global head movement was eliminated in four fragments but still slightly present in two fragments.

**Conclusion:** The best results were obtained by the non-rigid image registration method mesh-based warping using landmarks placed on the vertebrae and the jaw, and mesh-based warping using segmentation. The results of mesh-based warping using segmentation differed intersubjectively, presumably caused by segmenting different anatomical structures. We concluded that the best method to nullify the global head movement is the mesh-based warping method using segmentation. The segmented areas must correspond with the same regions of interest as the placed landmarks: the vertebrae, the upper and lower jaw, the back of the head, the orbit and the coin on the throat must be segmented and uniformly spread across the whole frame.

**Keywords:** videofluoroscopy; dysphagia; non-rigid image registration; global head movement.



# Table of Contents

<b>1</b>	<b>Reading Index</b>	<b>1</b>
<b>2</b>	<b>Introduction</b>	<b>3</b>
<b>3</b>	<b>Background</b>	<b>5</b>
3.1	Dysphagia and oncology . . . . .	5
3.2	Anatomical Background . . . . .	5
3.2.1	Swallowing process . . . . .	5
3.2.2	Videofluoroscopy (VFS) . . . . .	8
3.2.3	Treatment of head and neck cancer . . . . .	8
3.3	Technical background . . . . .	9
3.3.1	Medical image registration . . . . .	9
3.3.2	Similarity measures . . . . .	9
3.3.3	Landmarks . . . . .	10
3.3.4	Image segmentation . . . . .	10
3.3.5	Transformation models . . . . .	11
3.3.6	Image morphing . . . . .	12
<b>4</b>	<b>Problem and Goal</b>	<b>15</b>
4.1	Problem definition . . . . .	15
4.2	Importance of research . . . . .	16
4.3	Aim . . . . .	16
4.4	Research question . . . . .	17
4.5	Hypothesis . . . . .	17
<b>5</b>	<b>Method</b>	<b>19</b>
5.1	Research Methodology . . . . .	19
5.2	Fragment and frame selection . . . . .	19
5.2.1	Fragment . . . . .	19
5.2.2	Frame selection . . . . .	20
5.2.3	Contrast and brightness . . . . .	20
5.3	Landmarks . . . . .	21
5.4	Non-rigid transformation models in MATLAB® . . . . .	22
5.4.1	Demon registration . . . . .	22
5.4.2	Mesh-based warping . . . . .	22
5.5	Segmentation of anatomical structures . . . . .	23
5.6	Obtaining coordinates of the segments . . . . .	24
5.7	Comparing the results of mesh-based warping . . . . .	25
5.8	The absolute values of the difference of consecutive frames . . . . .	26
<b>6</b>	<b>Results</b>	<b>27</b>
6.1	Contrast and brightness . . . . .	27
6.2	Non-rigid methods . . . . .	27
6.2.1	Demons algorithm . . . . .	27
6.2.2	Mesh-based warping with landmarks . . . . .	28
6.2.3	Mesh-based warping with segments . . . . .	29

---

6.2.4	The absolute values of the difference of consecutive frames . . . . .	30
<b>7</b>	<b>Discussion</b>	<b>37</b>
7.1	Fragment and frame selection . . . . .	37
7.2	Landmarks . . . . .	37
7.3	Non-rigid registration methods in MATLAB <sup>®</sup> . . . . .	38
7.3.1	Demons algorithm . . . . .	38
7.3.2	Mesh-based warping . . . . .	38
7.4	Rigid registration methods in MATLAB <sup>®</sup> . . . . .	38
7.5	Segmentation . . . . .	38
7.6	‘ut_contourfft’ . . . . .	39
7.7	Validation of the obtained results . . . . .	39
<b>8</b>	<b>Conclusion</b>	<b>41</b>
8.1	Non-rigid image registration method . . . . .	41
8.2	Elimination of the global head movement . . . . .	41
8.3	Accuracy and reproducibility of the VFS assessment . . . . .	42
<b>9</b>	<b>Recommendations</b>	<b>43</b>
<b>10</b>	<b>Acknowledgements</b>	<b>45</b>
	<b>References</b>	<b>46</b>
<b>A</b>	<b>Results</b>	<b>51</b>
<b>B</b>	<b>Mesh-based warping script</b>	<b>55</b>
<b>C</b>	<b>Demon registration script</b>	<b>59</b>

# Reading Index

---

In the second chapter, a brief introduction of our research is given.

Subsequently, chapter three provides background information, containing information about dysphagia and different techniques that were used to carry out our research.

The fourth chapter includes our problem definition and the pursued goal.

In chapter five, the methods we used to answer our research question are described, such as the editing of the videofluoroscopy videos.

The obtained results are displayed in the sixth chapter.

The seventh chapter discusses our conducted research, followed by our conclusions in chapter eight.

At last, chapter nine, gives an impression of future research that could be done in addition to our study.



# Introduction

---

THE aim of this study was to establish a more objective, reproducible and accurate assessment of the swallowing function of patients with head and neck cancer. From a clinical perspective, an important prerequisite was that the developed method used to analyse fluoroscopic videos would by any means be time-saving. One of the requirements was that the global head movement of the subjects should be eliminated, without disturbing the representation of the swallowing process in the video. This was accomplished by editing the videos, using several non-rigid image registration methods. Based on the edited videos, a conclusion was made about which technique fulfilled the requirements and eliminated the global head movement most properly.



# Background

---

## 3.1 Dysphagia and oncology

IN 2010 6.9 per 1000 patients in the Netherlands suffered from dysphagia [1]. Dysphagia is the medical term for problems with swallowing, which includes an obstructed transport of fluids, solids or both, from the pharynx to the stomach. There are two types of dysphagia: oropharyngeal dysphagia and oesophageal dysphagia. Oropharyngeal dysphagia means that the patient has difficulties with emptying the oropharynx in the oesophagus. Oesophageal dysphagia is defined as problems with food passing through the oesophagus. This study focusses on the oropharyngeal dysphagia. Dysphagia can lead to oral secretion, malnutrition, dehydration or tracheal aspiration of the ingested food. Recurrent aspiration, also known as food or liquid entering the trachea, can lead to chronic pulmonary disease [2, 3]. Besides swallowing problems, dysphagic patients might also have an affected speech. Speech is one of the most powerful social interaction tools and loss of speech is often associated with severe functional and psychosocial problems, resulting in a poor quality of life [4].

Dysphagia can have a neurological or muscular cause and it can be due to a motility disorder or a mechanical obstruction. Examples of these causes are Parkinson's disease, multiple sclerosis, cerebrovascular accident and myasthenia gravis. Medication, radiotherapy and surgery can also cause dysphagia. Patients with head and neck cancer, who underwent chemoradiation or surgery, have a 50.6% prevalence of dysphagia [5]. Head and neck cancer covers oral, oropharyngeal, hypopharyngeal, nasopharyngeal, pharyngeal, laryngeal, nasal cavity, paranasal sinus and salivary gland cancer [6]. In the Netherlands every year about 17.54 out of 1000 people are diagnosed with oral and oropharyngeal cancer [7], which makes head and neck cancer to be the seventh most common cancer among men and ninth among women [8].

## 3.2 Anatomical Background

### 3.2.1 Swallowing process

Healthy people swallow between 800 and 2400 times a day in order to transport food from the oral cavity (mouth), through the pharynx (throat), to the stomach [4]. Swallowing is a complicated process using more than 25 muscles and is innervated through at least five different cranial nerves (CN.): nervus trigeminus (V), nervus facialis (VII), nervus glossopharyngeus (IX), nervus vagus (X) and the nervus hypoglossus (XII). The normal swallowing process consists of four phases: oral preparation, food transport in the oral region, food transport in the pharyngeal region and food transport in the oesophageal region [9, 10]. Shown below, in Figure 3.1, is the anatomy of the oral cavity, oropharyngeal region, nasopharynx, pharynx and larynx, where the swallowing process takes place. This region includes e.g. the tongue, the lips, cheeks, the vocal cords, the hyoid bone and the epiglottis.

#### **The first phase: the oral preparation phase**

In the first phase, the oral preparation phase, food is moistened with saliva, masticated and a bolus is formed. Then the food bolus is positioned on the tongue for transport, see Figure 3.2 below. In this

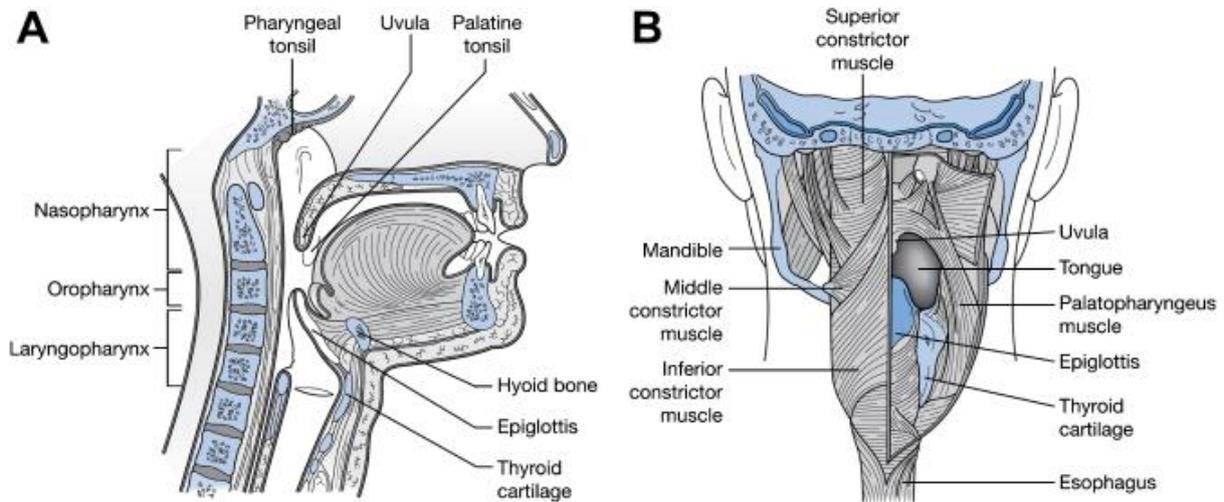


Figure 3.1: Anatomy of the structures involved in the swallowing process [11].

phase the nervus, abbreviated as n., trigeminus, innervates the lower jaw. The n. facialis innervates the lips, the n. vagus, n. accessorius and n. trigeminus innervate the movement of the upper roof of the oral cavity and finally the n. hypoglossus innervates the tongue [12].

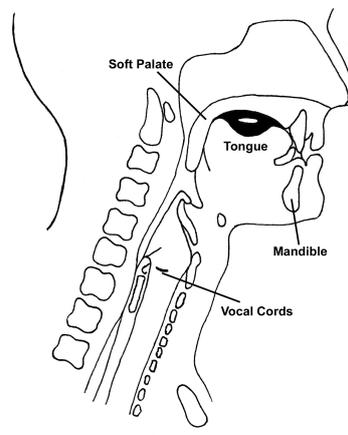


Figure 3.2: Oral preparation phase [13].

### The second phase: transportation in the oral region

In the second phase, food is transported from the oral cavity to the oropharynx, which takes only one second. The tongue pushes the soft bolus to the oropharynx, as seen in Figure 3.4. At the same time the musculus, abbreviated as m., obicularis oris and the m. zygomaticus seals the lips, and the m. buccinator moves the tongue. Also the uvula, the posterior edge of the velum seen in Figure 3.1, elevates in order to prevent food or liquid entering the nasopharynx. The oral phase ends when the bolus passes the beginning of the throat and touches the posterior wall of the pharynx. In this phase the n. hypoglossus innervates the tongue movement and with the n. vagus it innervates the floor of the oral cavity [12].

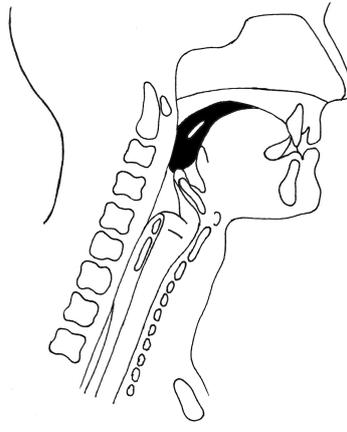


Figure 3.3: Oral transportation phase [13].

### The third phase: transportation in the laryngopharyngeal region

The third phase (see Figure ??) describes the bolus transport in the laryngopharyngeal region. When the bolus reaches the oropharynx, the swallowing process becomes an automatic process, instead of voluntary, and breathing stops temporarily. Due to elevation of the hyoid bone, the epiglottis and the vocal cords enclose to prevent aspiration. The velum or soft palate is tensed and elevated at this phase, which closes the uvula to prevent food or liquid entering the nasopharynx. To transport the bolus into the oesophagus, a pressure is created by converging the base of the tongue and the walls of the throat. In this phase the n. trigeminus, n. glossopharyngeal, n. vagus and n. accessorius innervate the movement of the upper roof of the oral cavity. The n. trigeminus, n. facialis, n. glossopharyngeus, n. vagus, n. accessorius and n. hypoglossus innervate the pharyngeal and laryngeal movement. The n. glossopharyngeus also innervates sensory input [12].

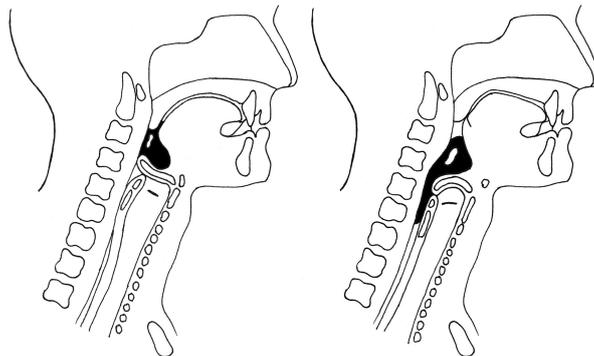


Figure 3.4: Food transportation in the laryngopharyngeal region [13].

### The fourth phase: transportation in the oesophageal region

In the last phase (see Figure 3.5), the bolus is transported through the oesophageal region, which lasts the longest (8-10 seconds). When the upper oesophageal sphincter opens, the bolus is transported downwards by squeezing actions of the throat muscles. When the lower oesophageal sphincter opens, the bolus can enter the stomach. This phase is innervated only by the n. vagus [12].

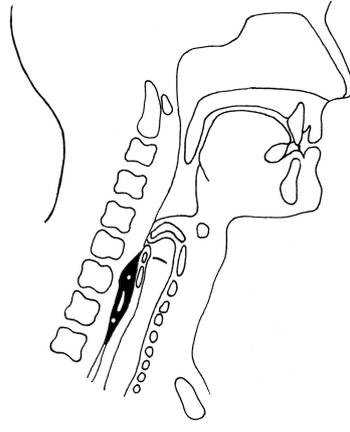


Figure 3.5: Food transport in the oesophageal region [13].

### 3.2.2 Videofluoroscopy (VFS)

Swallowing physiology varies as the bolus type changes. Parameters that need to be taken into consideration are quantity, consistency, texture and taste [14]. To evaluate the swallowing process, different methods are available: physical examination, rating scales, manometry, scintigraphy and image based evaluation (e.g. videofluoroscopy, endoscopy and cine MRI) [15, 16]. To analyse the patient's ability to swallow, videofluoroscopy is used most often. VFS is a real-time examination of the swallowing function of patients, which is observed with fluoroscopy: a technique that uses x-ray imaging. Different concerns can be observed in the VFS, such as the occurrence of aspiration and/or penetration, effectiveness of the different phases, spilling and amount of residu after the swallow [16].

When patients are not able to swallow normally, VFS helps to detect to what extent contrast medium is swallowed correctly or enters the trachea. During the VFS examination, patients swallow different amounts of a contrast medium, such as iodine, as a liquid, semi-liquid or solid bolus. X-ray beams and a contrast medium enable screening of the oral and oropharyngeal area. The safety, effectiveness and any evidence of aspiration is examined, by a diagnostic radiographer along with a speech-language pathologist (SLP) [17]. The current method for evaluation is time-consuming. First a SLP carries out the videofluoroscopy while another SLP takes notes. Hereafter, they evaluate the swallowing process by looking at the video and finally further treatment is discussed with a head and neck surgeon. The videos will be made according to the Logemann Protocol [13]. This protocol provides information about transport times, driving pressure, the opening of the oesophagus, penetration/aspiration and residues. Eventually the videos will be evaluated frame by frame (25 frames/minute). If aspiration already has occurred before treatment, adaptation of nutrition is needed before continuation of treatment, because these patients are at risk for developing an aspiration pneumonia [18].

### 3.2.3 Treatment of head and neck cancer

Before patients with advanced head and neck cancer undergo treatment, they often suffer from compromised swallowing, caused by the tumour. After treatment these functions may be even more compromised due to the treatment. There are different treatments and combinations of treatments possible: surgery, radiation therapy, chemoradiation therapy, surgery with postoperative (chemo)radiation therapy, photodynamic therapy (PDT) and laser therapy [6, 16]. The amount of functional loss after treatment depends on the location and extension of the tumour and the treatment modality that is used. For advanced tumours, the first treatment of choice is concurrent

chemoradiation (also called organ-sparing therapy). However, if a surgical resection leads to an unacceptable loss of function, a tumour is diagnosed as functional inoperable. In these circumstances (chemo)radiation serves as an alternative, with the risk of developing fibrosis and xerostomia (dry mouth). The side effects can be so severe that permanent tube feeding is necessary, which is the case in 13% to 60% of the people who undergo chemoradiation for oral or oropharyngeal cancer [4, 19, 20]. Intensity-modulated radiotherapy (IMRT), a complex computerised treatment planning, is currently used more often. It reduces the dose that the healthy tissue will be exposed to, which results in lower long-term xerostomia rates [4, 21]. Prediction of (unacceptable) functional losses after surgery and/or chemoradiation is subjective and digital biomechanical modelling of the oral cavity and oropharynx to perform virtual therapy, might contribute to an objective virtual assessment of swallowing and speech disorders [4].

## 3.3 Technical background

### 3.3.1 Medical image registration

In some imaging techniques, such as MRI, it can be desirable to match the exact positions of anatomical features between images, so that they can be properly compared. Therefore, the aim of image registration is to find corresponding features in two or more frames and provide an accurate match of the two frames [22]. There are two types of image registration: rigid registration and non-rigid registration. Rigid registration applies spatial transformations to a frame, such as rotation, translation and scaling. This is commonly used for corrections of small movements of rigid parts of the body like bone, but cannot be used for dislocation of soft tissue due to non-homogeneous movement [23]. Non-rigid registration is applied when non-linear transformations, such as local stretching, are needed to properly match two frames. There are three common applications for non-rigid registration: reducing artefacts in imaging systems, reducing image distortion in soft tissues and comparing inter-subject images. Another application of non-rigid registration is comparing intra-subject images. VFS is a technique that also compares intra-subject images. A registration algorithm can be divided into three components:

1. A similarity measure: how well do two images match?
2. A transformation model: how can the target image be matched to the source image? This model defines parameters involved in the transformation.
3. An optimization process: it varies some of these parameters of the transformation model to assure maximum correspondence [24].

### 3.3.2 Similarity measures

There is a distinction between geometry-based and intensity-based approaches of non-rigid registration. The geometric approach uses surfaces or landmarks, often on functionally important anatomical structures, which are matched in each frame. This use of structural information can define the displacement of the anatomical features from image to image and makes sure that it is anatomically valid. After matching landmarks in different frames, interpolation is necessary to match the complete frames to each other [24]. Using interpolation, one can calculate missing data points based on available information, in this case the landmarks [25]. The intensity-based approach uses mathematical or statistical criteria to match intensity-patterns in two frames. Intensity-based relationships only match intensity and do not use similarity between anatomical structures. Therefore there are also geometric-intensity combined similarity measures, which takes both, anatomical similarities and intensity-based similarities, into account. There is a collection of methods available to measure similarity between the source and the target, categorised in application for monomodal or multimodal image registration (e.g. when comparing a MRI-scan with a CT-scan) [24].

### 3.3.3 Landmarks

As described above, a geometry-based similarity measurement takes anatomical similarities into account by using a set of control points or landmarks in two images. There are two types of control points used in image registration: fiducial markers and landmarks. Fiducial markers are attached to a patient's body during a scan. It can be a simple object, like a coin, and is clearly visible on the scan or image. Therefore, they can be easily used as anatomical point references. Fiducial markers are mostly used for rigid image registration and multi-modality registrations. Disadvantages of using fiducial markers are that the markers can be moved during the examination when the skin is displaced and that they can only be attached to anatomical points on the exterior of the body citehaidekker.

Instead of external fiducial markers, anatomical landmarks can be used. These are point references on anatomical features that can be manually placed on an image after the examination. One of the advantages is that it is less time-consuming to make the scans or images, because the markers are not placed on the patient during the examination. Also, anatomical positions inside the body can be used for image registration and the researcher can vary the positions of landmarks [23].

### 3.3.4 Image segmentation

Another way of creating landmarks is using segmentation. The boundary of a segmented region can be converted into landmarks. The definition of image segmentation is the subdivision of distinct regions in an image. The regions have similar properties, such as color or contrast, and represent different tissues or different anatomical features [26]. The most important goal is to divide the image in regions to study the separate anatomical structures or identify regions of interest, for example tumours [22,27]. Segmented image are often useful as input for automated image analysis. One of the conditions the region has to meet, is that it is uniform: it has to consist of only one type of tissue. Another condition is that the boundaries of the regions are well-defined [26]. Automated segmentation, creating the region of interest as a segment automatically, is frequently employed by radiologists in diagnosis and treatment planning. There are various automated segmentation algorithms, specific for the imaging modality or part of the body it is used for [27].

### 3.3.5 Transformation models

The definition of a transformation model is an algorithm that defines how the target image is transformed to match the source image. A transformation model has two functions:

1. For those aspects of an image where there is no data available, the transformation model transforms the image using interpolation.
2. It controls the transformation of image features, relative to each other with the aim of achieving the greatest correspondence [24].

Rigid or affine transformation uses six to twelve parameters, applied to a vector, that locate the matching coordinates in a second image. Rotation and translation in three directions can be described by six parameters and six other parameters describe scaling and shears. Non-rigid registration uses other transformations to match frames. These vary from local variations, described with a few parameters in a transformation model, to complex dense displacement fields, which give information about each voxel. There are several types of non-rigid registration transformation models, which include the following, but are not limited to:

- Radial basis functions (RBFs)
- Physical continuum models
- Finite element models
- Image morphing techniques

Which transformation model is most suitable, depends entirely on the application.

#### Radial basis functions

Radial basis functions are radially symmetric functions, in which the function value at each point depends on the distance of the point to the origin [28]. The type of radial basis function determines the characteristics of the transformation model, like the locality, the solvability and the efficiency. Commonly used radial basis functions are thin-plate splines, (inverse) multiquadrics and the Gaussian [29]. The group of spline algorithms, including thin-plate splines and B-splines, has been used in medical image registration for many years and in various applications. Spline functions are essentially multiple polynomials, joined together at points called knots. The higher the order of the polynomials and the more knots, the more flexible the spline [30]. Thin-plate splines, or surface splines, are commonly used for non-rigid (elastic) image registration [31]. It is an analytic solution that produces a surface over a scattered dataset using interpolation or approximation [32]. A thin-plate spline can be imagined as a plate of thin steel that is fixed to certain constraint points, or landmarks. When these landmarks are displaced, the bending of the steel plate costs energy, defined as the bending energy. Thin-plate splines bend in such a way that the bending energy is minimized. The displacement of a landmark in one corner of the plate will have a global effect on the entire plate [33].

#### Physical continuum models

Physical continuum models can be divided in elastic models, visco-elastic models, the demons algorithm and more [24, 34]. A physical continuum model is based on the idea that the transformed image is embedded in a three-dimensional deformable elastic or viscous fluid medium. Due to a displacement field, the medium and eventually the target image will be reconfigured [35]. However, this model, when based on an intensity similarity measurement, compares differences pixel-by-pixel and therefore stable anatomical structures, such as bone, are not entirely taken into account. Therefore, physical continuum

models are mostly used to transform completely soft tissue, like brain tissue. When a physical continuum model is based on a geometrical similarity measurement, soft tissue can be transformed using stable structures [36].

### Finite Element Models

The finite element model controls localized deformation. It divides an image into cells and describes the physical nature of the cell, e.g. “elastic” for soft tissues, and “rigid” for bone. It is most commonly used in applications where there are biomechanical restrictions and mechanical forces are being measured [37].

### 3.3.6 Image morphing

Image morphing is an image processing technique used for the metamorphosis from one image to another. The simplest method of transforming one image into another is to cross-dissolve between them. This implies that the color of each pixel is interpolated over time from the first image to the corresponding second image value. However, this is not very effective for the actual metamorphosis, because two images can have a slightly different shape and on that account will not morph well. Therefore, a morphing process which combines cross-dissolving with warping methods would give better results. The morph process must then consist of a warping stage before cross-dissolving, in order to align the two images [38].

In case of two images being morphed, the first image is gradually distorted and faded out. The second image gradually fades in. The combined frame contains fifty percent of both images. There are many implementations of image morphing. The algorithms used for morphing are based on mesh warping, field morphing, radial basis functions, thin plate splines, energy minimization, and multilevel free-form deformations [39].

### Mesh-based warping

Mesh morphing is a feature-based metamorphosis technique that uses two meshes: a source mesh, from the first image and a target mesh, from the second image. The algorithm aligns a rectangular grid of points, or mesh, to key feature locations of the images. Those key feature locations are determined in advance [40]. The goal is the construction of a correspondence map between the source and target, using bicubic spline interpolation to compute warp functions [41]. In the warping phase, a destination grid is calculated with coordinates that are weighted averages of the corresponding coordinates of the source grids. For the warping algorithm, the coordinates of the source and destination grids are fitted with piecewise continuous mapping functions, used for calculating the correspondences between the source and destination images. First an intermediate destination image is generated in which every row of the image is warped separately from the other rows. Second the final destination image is created by warping every column of the intermediate image [40].

Mesh-based warping uses bicubic spline interpolation. Bicubic splines are a prolongation of cubic splines, that is to say, a cubic spline can interpolate one-dimensional data points, whereas bicubic splines interpolate data points on a 2D-grid. A cubic spline is a piecewise function that consists of a certain number of third degree polynomials:

$$S(x) = \begin{cases} s_1(x), & \text{if } x_1 \leq x < x_2, \\ s_2(x), & \text{if } x_2 \leq x < x_3, \\ \dots & \\ s_{n-1}(x) & \text{if } x_{n-1} \leq x < x_n. \end{cases} \quad (3.1)$$

$s_i(x)$  is a third degree polynomial:

$$s_i(x) = a_i(x - x_i)^3 + b_i(x - x_i)^2 + c_i(x - x_i) + d_i \quad (3.2)$$

A cubic spline interpolant should have the following characteristics:

- The piecewise function  $S(x)$  interpolates all data points and is continuous on the interval  $[x_1, x_n]$ .
- The first and second derivatives of  $S(x)$  are continuous on the intervals, which makes a cubic spline a smooth interpolating function [42].



# Problem and Goal

---

## 4.1 Problem definition

NOWADAYS, several speech-language pathologists (SLPs) use the Logemann protocol to evaluate the different swallowing phases in videofluoroscopy recordings [13]. These recordings are manually analysed, often frame by frame in slow motion, and later discussed with a head and neck surgeon. Due to the fact that patients must swallow multiple times with different consistencies and amounts, there are a lot of frames that need to be assessed, which is very time-consuming. Another issue concerning videofluoroscopy is that it is manually assessed and therefore subjective. Research has shown that the assessment of videofluoroscopy differs per expert [43]. Researchers and SLPs, working with patients who suffer from dysphagia, have tried to make the evaluation more clinically relevant, less time-consuming and less subjective. Examples of these attempts are the Dysphagia Outcome and Severity Scale and the Modified Barium Swallow Impairment Profile. Still, these measurements must be evaluated manually [44, 45]. It would be desirable to have an automatic assessment of the fluoroscopy videos, to get a more accurate evaluation of the swallowing function. The automatic or computerized assessment of the VFS would enable doctors to have a quantitative analysis with the use of physical parameters, such as the estimation of volume and flow of the bolus. The reproducibility of the VFS assessments will also increase if the assessment becomes automatic. However, the question remains whether automation will lead to a more qualitative assessment. If this is the case, it would be a step in the right direction to establish a better prediction of the functional outcome for patients after treatment. In the future, this will be useful for the consideration of a suitable treatment for the patient, looking at what will provide the best quality of life. In the current situation there is ambiguity about the quality of life after a treatment, which causes patients to have an uncertain vision of their future. Doctors and surgeons cannot entirely predict to what extent swallowing function loss will occur after surgery or chemoradiation. Nowadays, they use their knowledge and experience to predict which general functional outcomes might occur after treatment. This makes it difficult to predict with a hundred percent certainty whether a tumour is functional inoperable [46]. Therefore, properly informing patients about personal consequences after treatment is impossible.

The most important issue at the moment is the global head movement of the patient during the analysis of the VFS. Previous research has shown that it is impossible to (semi-)automatically follow the flow of the bolus when the patient's head is moving [47]. Patients with dysphagia who have problems with swallowing can swallow in a strained way to improve their swallowing function. This may result in extra movements of the head, which has a negative influence on the assessment of the swallowing function [18].

At this moment, the analysis of the videofluoroscopy is the golden standard for an objective analysis of the swallowing function. However, analysis must still be done manually, which is time-consuming. Also the videofluoroscopy frames cannot give detailed information about e.g. the amount of residue after swallowing and/or estimation of volume and flow, which would provide detailed and important information of the swallowing function. It can be concluded that more research is needed.

## 4.2 Importance of research

Besides head and neck cancer, there are many more disorders leading to dysphagia, such as a stroke, which is a neurological disorder. Of all people who have suffered from a stroke, 50-70 % will suffer from dysphagia [48]. Another big concern is aspiration pneumonia, a consequence of dysphagia. This is a leading cause of death among the elderly and has been reported as a growing cause of hospital admissions among this population [48]. Determination of the quantity and localisation of the contrast medium would therefore be helpful to predict the swallowing process and risks of lung disease [2]. It is difficult to determine how much contrast medium remains in the oral cavity, oropharynx or hypopharynx, for example in the sinus piriformis or above the epiglottis [18]. Early diagnosis and intervention to minimize the risks of oropharyngeal dysphagia are considered to be essential in dysphagia management [49]. Therefore, it is of great clinical importance that the assessment of VFS improves and leads to a more objective and accurate evaluation of the swallowing process.

Looking at the social relevance, automatic processing of videofluoroscopy could lead to a reduction of healthcare costs. This reduction could be accomplished in at least two ways. First of all, reducing the staff will lower costs, since the VFS analysis time will be shortened. Secondly, automation of the VFS will indirectly lead to lower healthcare costs due to a more accurate diagnosis. A better prediction of risks could result in more effective treatment. Treatment of severe pneumonia costs at least 25.000 dollar per treatment, while one VFS procedure costs at most 500 dollar [50]. Literature shows that VFS is already more cost effective than other diagnostic tools and a better prediction will eventually reduce the amount of people with pneumonia caused by dysphagia [50].

Looking at the scientific relevance, automatic assessment of the VFS will contribute to the research of the Virtual Therapy Group ([www.virtualtherapy.nl](http://www.virtualtherapy.nl)). Their futuristic mission is to construct a digital model of every cancer patient, which will predict the expected functional outcomes [51].

As discussed above, there are a lot of causes for dysphagia and the consequence can be severe. Additionally, an improvement of diagnosing dysphagia could reduce health-care costs. Last, an automatic assessment of the VFS will contribute to future research about virtual modelling of patients. Concluding, this outlines the great importance of this research.

## 4.3 Aim

As mentioned, our research is part of a bigger project of the Antoni van Leeuwenhoek hospital in collaboration with the University of Twente. This bigger project has three goals. First of all to improve the objectivity, the reproducibility and the accuracy of the assessment of swallowing function based on VFS. Second, reducing the time needed for VFS assessment. The third goal is to enable patient-specific modelling of the swallowing function. These improvements would contribute to the main purpose of the Virtual Therapy Group project: to be able to predict the function loss, for head and neck cancer patients, after a certain treatment with the use of video analysis of the swallowing process and a virtual model. Hereby, the best treatment could be chosen in advance and will be patient-specific.

As described in the problem definition, it is essential to eliminate the global head movement in order to accomplish a (semi-)automatic assessment of the VFS. Therefore, the aim of our research is to eliminate the global head movement of the patient during the VFS, in order to assess the swallowing process more accurately. This can be the first step towards a more objective and accurate VFS analysis. If this research leads to positive results, a follow-up research could be done to accomplish the automatic evaluation of the swallowing function.

## 4.4 Research question

We formulated the following research question:

*To what extent can non-rigid image registrations be used to nullify the global movement of healthy subjects during the swallowing process and eventually result in a more objective assessment of the videofluoroscopy?*

To be able to answer the main question, we formulated two sub-questions:

- *Which non-rigid image registration method is best to nullify the global movement of the healthy subject's head?*
- *Does the correction of the global movement lead to a more accurate and therefore more reproducible assessment of the videofluoroscopy?*

## 4.5 Hypothesis

We expect to be able to use methods of non-rigid image registrations implemented in MATLAB<sup>®</sup> [52] in order to eliminate the movement of the patient's head, the so-called global movement of the patient. Non-rigid image registrations can use non-linear transformations to register the frames of one fluoroscopy video. Using this approach, the patient's head remains stationary, which enables the analyser to focus on the swallowing process, the local movement of the patient. Based on literature research, we think a geometric approach as similarity measurement will be most suitable to match frames, because earlier research had shown that the intensity-based method was not suitable [47]. With a geometry-based approach we can apply anatomical landmarks on characteristics of the VFS videos and match them. To match the frames with anatomical landmarks, we think a radial basis function using splines or a physical continuum model as transformation model will be most suitable to eliminate the global movement. The transformation model has to be able to use a geometric similarity measure.

Thus our prediction is that non-rigid registration could be used to freeze global movement relative to local movement. This will provide an improved monitoring of the bolus and will lead to a more accurate and therefore more reproducible assessment of the videofluoroscopy.



## 5.1 Research Methodology

IN order to achieve our aim and answer our research question, we have set up a retrospective research: existing fluoroscopy videos of eight healthy subjects have been analysed. We have worked with MATLAB<sup>®</sup> to analyse the videos and to nullify the global movement of the subject's head relative to the local swallowing movement. Windows Movie Maker and MPEG streamclip are used to cut the fluoroscopy videos into fragments. With a geometric similarity measure, we have matched the different frames of one fluoroscopy video. After matching the frames, we transformed the videos with non-rigid image registration methods. The non-rigid methods we used were the demons algorithm, a physical continuum model, and mesh-based warping, based on bicubic spline interpolation. After editing, the movement of the subject's head has been compared to the movement of the original videos.

In the background we described that a medical image registration algorithm contains a similarity measure, a transformation model and an optimization process. We only focussed on the similarity measure and transformation model.

## 5.2 Fragment and frame selection

### 5.2.1 Fragment

We have received videos of eight volunteers from the Antoni van Leeuwenhoek hospital in AVI-format. These videos were made according to the adjusted Logemann protocol. The videos contained multiple swallowing processes with different amounts and consistencies of contrast medium. First we renamed the videos by assigning numbers to the subjects from one to eight, in order to make the subjects anonymous. We have not analysed all the videos, but made a selection by watching and grading the videos in advance. We selected videos with the amount and consistency that showed the most visible movement of the subject's head for analysis. Those videos would show the biggest contrast between the unedited and edited frames. We used the four videos that showed the biggest head movement, because when the editing succeeds in these videos, it will probably work on smaller movements.

In order to make an appropriate selection, we analysed the movement per subject per consistency and gave a score from 0 to 3: 0 implying no movement of the head and 3 a lot of movement. Every video has been watched by two people and has been given a score for the amount of movement, see Table 5.1. The swallowing processes with the highest number in general were chosen to edit. The grading showed that the swallowing process with gingercake had the biggest movement. However, we wanted to analyse two similar amounts with different consistencies, in order to eliminate a difference in movement due to different amounts. Meaning, we wanted the fragments thick and thin with the same amount of contrast medium. According to the table both, 3cc and 5cc, were suitable, but eventually we chose 5cc because of the assumption that a larger amount of contrast medium results in bigger head movements. We named the videos after the number of the subject and the consistency: 5cc thin became video A of the subject

and 5cc thick became video B. For example, 5cc thin fluoroscopy video of subject number 2 will be named video number 2A.

Table 5.1: Assessment of global head movement in swallowing videos

Number of subject	1cc thin	3cc thin	5cc thin	10cc thin	3cc thick	5cc thick	gingercake
Subject nr. 1	1	1	1	0.5	1	0.5	1
	1	0.5	1	1	1	0.5	1
Subject nr. 2	0	2	1.5	3	2.5	3	1.5
	0	2	1	3	2.5	2	1
Subject nr. 3	0	1	2	1	1.5	2	1.5
	0	1	1	1	2	1.5	1.5
Subject nr. 4	2	1	1	1	3	2	2
	1	1	1	1	2	2	3
Subject nr. 5	0	0	0	0	0	0	2
	0	0	0	1	1	2	3
Subject nr. 6	0	1	1	0	1	1	-
	0	0	0	0	1	1	-
Subject nr. 7	1	0	1	1	2	2	3
	0	0	0	1	2	2	2
Subject nr. 8	0	0	0.5	0	0.5	0	1
	0	0	0.5	0	0.5	0	1
<b>Total</b>	<b>6</b>	<b>10.5</b>	<b>12.5</b>	<b>14.5</b>	<b>23.5</b>	<b>21.5</b>	<b>24.5</b>

We cut the original videos in videofragments with Windows Movie Maker and MPEG streamclip. In this way, one videofragment contained one swallowing process of 5cc thin or 5cc thick. Thus in total we had sixteen videofragments at our disposal for the analysis.

### 5.2.2 Frame selection

After the selection of video fragments has been made, we started the editing process. Our first step in selecting frames was to load the videos into MATLAB<sup>®</sup>. This program contains the necessary tools to edit the video fragments. We have chosen to edit and analyse five frames of each video fragment, because it would be too time-consuming to edit all the frames and since we want to freeze the movement of the subject's head, we needed five unedited frames that really showed movement of the head. Therefore, the five frames should be spread out within one swallowing process instead of following one another closely. We took into consideration that the chosen frames were focussed and that the bolus was visible.

### 5.2.3 Contrast and brightness

In order to select landmarks, the anatomical structures needed to be clearly visible. Increased contrast and brightness could help to define the lines between different anatomical structures. The adjustment of the contrast and brightness of an image can be done in MATLAB<sup>®</sup>. In order to do so, the image needs to be a grayscale image, using 'rgb2gray', and the precision needs to be increased. Increasing the precision is possible using 'im2double' [53]. Finally, two parameters which represent the contrast and brightness can be adjusted.

## 5.3 Landmarks

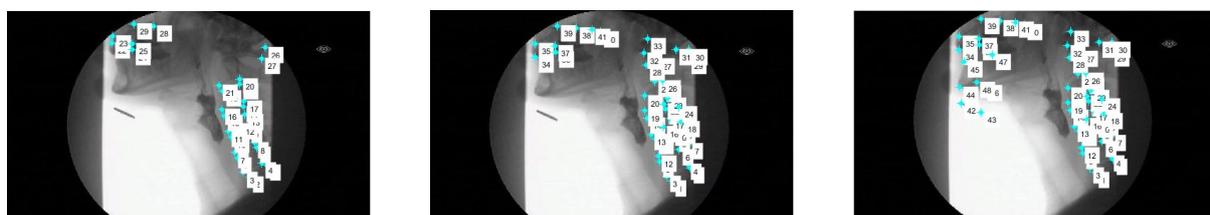
To analyse the displacement of the subject's head, we chose to work with landmarks on stable anatomical structures in the first place. Those landmarks were placed on structures that were not involved in the swallowing process: the vertebrae, upper jaw and orbit. In this way we wanted to prevent that the swallowing movement would be too much affected by the editing. We have chosen to place landmarks on two different subjects only, because it was not definite that the method would work and would be precise. We placed landmarks according to the following protocol:

- In the first place, no landmarks will be placed on non-stable anatomical structures involved in the swallowing process such as: lower jaw, throat, hyoid bone, oesophagus, etc.
- Vertebrae: A minimal amount of four vertebrae with landmarks, per vertebra four landmarks.
- Upper jaw: A minimal amount of four landmarks on the upper jaw, if tooth-implants visible, landmarks were placed on the tooth-implants.
- The orbit: if visible on the video, a minimal amount of two landmarks on the orbit.

To examine the influence of the amount of landmarks and the places the landmarks were placed on, we used three different ways of placing landmarks:

- A minimal amount of landmarks:  $< 30$  landmarks on the frame, see Figure 5.1a.
- A large amount of landmarks:  $> 30$  landmarks on the frame, see Figure 5.1b.
- A large amount of landmarks  $> 30$  landmarks, including landmarks on the moving structures as well, such as on the lower jaw and the coin on the throat, see Figure 5.1c.

We have chosen 'cpselect', a MATLAB<sup>®</sup> tool, to work with, in order to place the landmarks. 'Cpselect' is a graphical user interface, named 'Control Point Selection Tool', in which you can manually select control points in two related frames [54]. We labelled the control points or landmarks in the reference frame 'reference\_points' and in the moving frame 'moving\_points'. First, we placed the landmarks on the reference frame and the first moving frame. After that, we placed the same landmark sequence on the second moving frame and continued this process until all moving frames were completed. Due to the same landmark sequence in the reference and all the moving frames of one videofragment, the landmarks in different frames can be matched to one another.



(a) Selecting landmarks less than 30

(b) Selecting landmarks more than 30

(c) Selecting landmarks on the jaw

Figure 5.1: Selecting landmarks using 'cpselect'.

## 5.4 Non-rigid transformation models in Matlab<sup>®</sup>

Since rigid registration can only accomplish rotation and translation, we used non-rigid registrations to transform the fluoroscopy videos. In order to obtain the best results, we have tried two different non-rigid methods: Demon registration and mesh-based warping.

### 5.4.1 Demon registration

The demons algorithm is a physical continuum model, which can be used to align images. In MATLAB<sup>®</sup>, ‘imregdemons’ is a function that estimates the displacement field that aligns two 2D or 3D images. In this way, the displacement field of our different frames can be registered by comparing the reference and the moving frame. After registering all the frames of one video, the edited and unedited video can be compared.

We implemented the function `imregdemons` [55], in order to edit our frames. The demons function in MATLAB<sup>®</sup> is defined as followed:

```
[D,moving_reg] = imregdemons(moving,fixed),
[number of iterations], ‘AccumulatedFieldSmoothing’, (number of levels).
```

We used the first frame of each video as reference image and the consecutive frames as moving images. The number of iterations determines how much repetitions the function ‘imregdemons’ will compute. More iterations give a more accurate transformation, but it also takes longer. We want to edit the frames at the most suitable resolution, in order to get the best transformed frames [55]. The function requires the input images to be grayscale. Therefore we used the MATLAB<sup>®</sup> function ‘`rgb2gray`’. Since we already use black-white videos, the videoframes remained black-white.

The displacement is represented at each pixel location with a displacement vector. A displacement field is applied on the moving frame using linear interpolation to warp this frame. We plotted this warped frame next to the reference frame to show the difference. This was repeated for all the moving frames of one video. This resulted in transformed frames, based on the displacement field between the reference frame and one moving frame. See Appendix C for the entire script.

### 5.4.2 Mesh-based warping

Demons registration accomplishes a transformation based on a displacement field, but does not take anatomical structures into account. Therefore we decided to use a transformation model that uses landmarks. We used an existing mesh-based warping MATLAB<sup>®</sup> script, based on bicubic spline interpolation [56], see Appendix B. The mesh-based warping function is used to transform our moving frame to a reference frame using a matching set of landmarks. The total amount of landmarks in each frame has to be equal. The output of the warping is a transformed frame. From these transformed frames, we created videos to show our five consecutive frames that were all warped with reference to the first frame, the reference frame. Mesh-based warping is executed by creating an intermediate grid of the landmark-coordinates of both frames. Hereby, the value of variable alpha determines the weight both data sets have in creating this grid. This grid is used to warp both frames, reference and moving, to the intermediate grid. Because transformation of the moving frame is the aim, the reference frame should be dominant in creating the intermediate grid, which results into the desired transformation of the moving frame. When displaying the warped frame, the moving frames should be more visible than the reference image. To achieve this, we created a second variable beta that influences which frame is displayed more dominantly.

## 5.5 Segmentation of anatomical structures

In order to increase the precision of the non-rigid editing of the frames, we segmented anatomical structures. The segmentation provided more landmarks on both stable and moving structures compared to manually selecting landmarks. Stable structures are for example vertebrae, while the lower jawbone is a moving structure. The positions of the segments can be seen in the following figures: Figure 5.2a, shows a segmented vertebra; Figure 5.2b shows a segmented section of the lower jaw; Figure 5.3b, shows another segmented section of the lower jaw/chin; Figure 5.3a, shows a segmented upper jaw. We segmented frames of four subjects, instead of just two subjects used with the manual selection of landmarks, because segmentation was less time-consuming. There are various tools that can be used in MATLAB<sup>®</sup> for segmentation, such as ‘Impoly’, ‘Imfreehand’ and ‘ImageSegmenter’. These functions enable the user to select points or draw a line around the area of interest. We selected and tried the following functions:

- ‘Impoly’ creates a polygon using point selection. This polygon is draggable and resizable, which allows adjusting of the polygon and thereby increases the accuracy. The polygon will be a region of interest containing the selected points [57].
- ‘Imfreehand’ resembles the Impoly tool, aside from the fact that the region of interest is now selected by drawing a line around the area instead of manually selecting points. ‘Imfreehand’ provides a larger number of control points, which makes it more accurate. A disadvantage is that displacement and adaption is not possible when drawing a line with ‘Imfreehand’. This makes it difficult to achieve the same accuracy in comparison with ‘Impoly’ [57].
- ‘ImageSegmenter’ could be used in both ways, drawing a line and selecting points. This could be seen as cutting a region of interest out of the image which creates a segment.

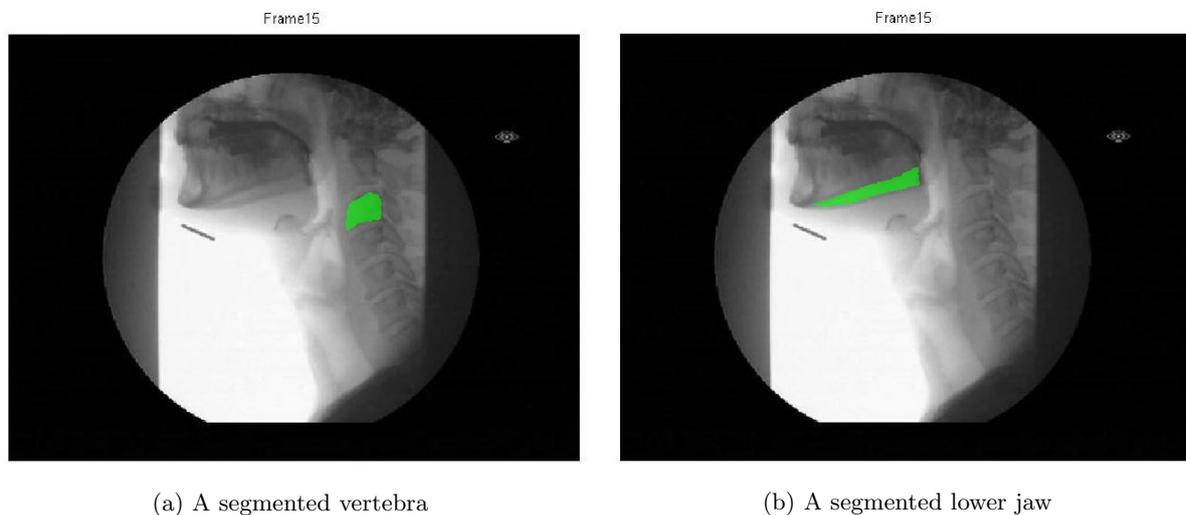


Figure 5.2: Selecting segments using ‘ImageSegmenter’.

The created segments of ‘ImageSegmenter’ consist of three different types of images: a mask, a BlackWhite (BW) and a maskedImage. These image types can be used in the mesh-based warping algorithm. This tool does not allow resizing and dragging of the selected line. Nevertheless, using the point selection of ‘ImageSegmenter’, a precise selection of the region of interest can be made [58]. Therefore, we segmented our structures with ‘ImageSegmenter’,

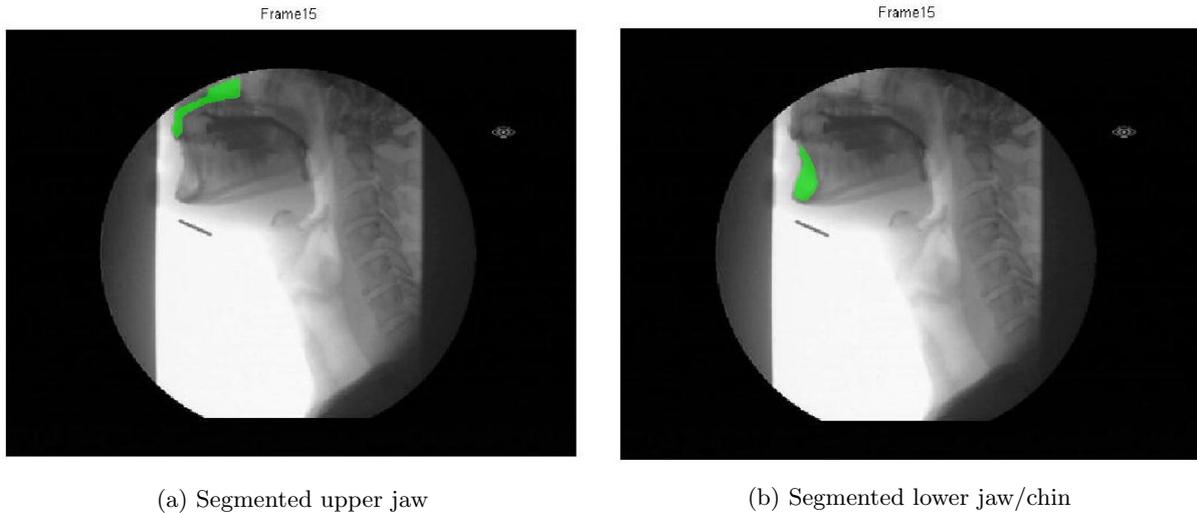


Figure 5.3: Selecting segments using ‘ImageSegmenter’.

To implement our landmark-data in our transformation model, all data should be saved into a matrix containing x- and y-coordinates. Data from ‘ImageSegmenter’ is saved as an image and not in data points. Only the BW, a segmented binary image, is used in our transformation model [53, 58].

## 5.6 Obtaining coordinates of the segments

To use the segmented parts in our transformation model, we needed to obtain the coordinates of the segments. By obtaining the coordinates, the segments can be processed into our non-rigid registration method mesh-based warping, where we compared the reference segments with the moving segments.

First we created a boundary of the BW-image of the segments using the MATLAB<sup>®</sup> function ‘bwboundaries’. After creating the boundary of the segments, we obtained an equal amount of coordinates per segment structure by editing the boundary with ‘ut\_contourfft’, which is a function used for shape or contour analysis [59]. This equal amount of coordinates per segment structure is needed to process the segments in our transformation model. ‘ut\_contourfft’ is a function that evaluates the closed boundaries of images or segments (BND) in the Fourier domain and resamples the contour.

```
[FDS,USPOLY] = ut_contourfft(BND);
```

USPOLY is a variable that represents the uniformly resampled boundary. Resampling results in a fixed set of points, specified by nfft, equally spaced along the running arc length of a contour. For example:

```
[~,USPOLY] = ut_contourfft(BND, nfft, 64);
```

‘ut\_contourfft’ also determines different measurements that are useful for establishing equal starting points. `Fdparm(5)` gives information about “the starting point expressed as a distance measured along the running arc length and with reference to the starting points of the normalized contour” and `fdparm(7)` is the outline of the selected contour [59]. In the example above, ‘ut\_contourfft’ resamples the boundary using 64 uniformly spaced points, but this does not automatically provide the same starting points of the boundaries. To compare two resampled data sets, we modified the boundary and

created corresponding starting points. This means that the starting points are in the same position of the contour, for example both on the left corner of the vertebra, see Figure 5.4. To determine the two starting points, we used the following function:

```
[~, ~, fdparm] = ut_contourfft(bnd, 'normfd');
u1 = fdparm(5)/fdparm(7);
```

After applying all of these steps, the starting points are shifted to a corresponding position using the following formula:

$$z(s + u) \leftrightarrow Z_k \exp\left(\frac{2\pi jku}{P}\right) \quad (5.1)$$

The data is returned to the time domain by using inverse fourier transformation. This makes it possible to save x- and y-coordinates and load this data into our transformation model. To facilitate loading, a matrix, with xy-coordinates of all segments of one frame, is made. These coordinates are loaded into our transformation model.

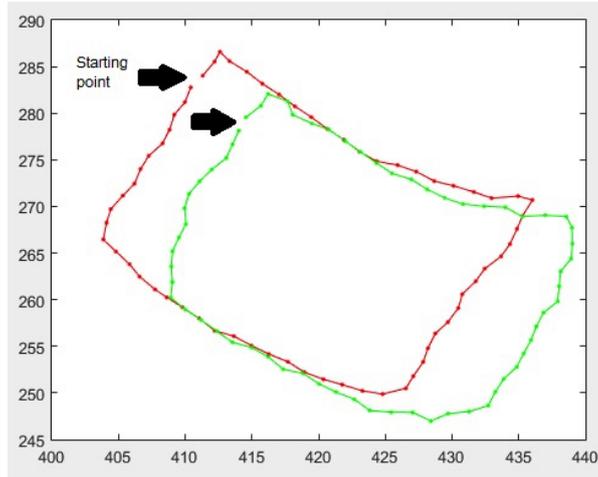


Figure 5.4: The uniformly resampled boundary of one of the vertebra. The starting points are shifted as described above.

## 5.7 Comparing the results of mesh-based warping

In order to analyse and compare the different transformation methods we used to edit the video frames, we have chosen to show the edited frames of one fragment in one table, which can be seen in the chapter results. This table shows the unedited selected frames of the videos, the frames edited using the different landmarks methods and the frames edited using segmentation. Each column in the table displays the comparison of the first frame, the reference frame, to one of the moving frames. The moving frames are frame number two to five. The comparison is made with 'imhistmatch', a function in MATLAB<sup>®</sup>, where you place one image on top of another. Both images get a different color: the reference frame will become purple and the moving frame will become green. The output frame shows the resemblance between the two images in grey and the differences are shown in either purple or green, depending on the reference or moving frame. An example of such an frame can be seen in Appendix A.

## 5.8 The absolute values of the difference of consecutive frames

After we obtained data from segmentation and mesh-based warping, we wanted to investigate whether our results are applicable for further research. In order to determine the influence of the amount of segments and frames, we segmented nine regions in ten frames instead of five frames with approximately six segments. The ‘ImageSegmenter’ tool was used once more to segment anatomical elements. The chosen structures were: the chin, lower jaw, the orbit, four vertebrae, the coin on the throat and a part of the back of the head. We used ‘mesh-based warping’ to transform our frames and combine them into a video. Hereafter, we applied the tool ‘imabsdiff’ [60], which determines the absolute difference between two frames. We compared every frame with the consecutive frame. The obtained frames with ‘imabsdiff’ were then made into a video wherein only the absolute difference between those frames was displayed.

## 6.1 Contrast and brightness

FIGURE 6.1a and Figure 6.1b show the same frame of one VFS video. Figure 6.1a is the original frame and Figure 6.1b the adjusted frame. Figure 6.1b is adjusted by changing the contrast and brightness settings. This results in more distinct bone structures, but the modified frame shows a more pixel-like image and the image is blurred. These results were not desirable and therefore not used in further research.

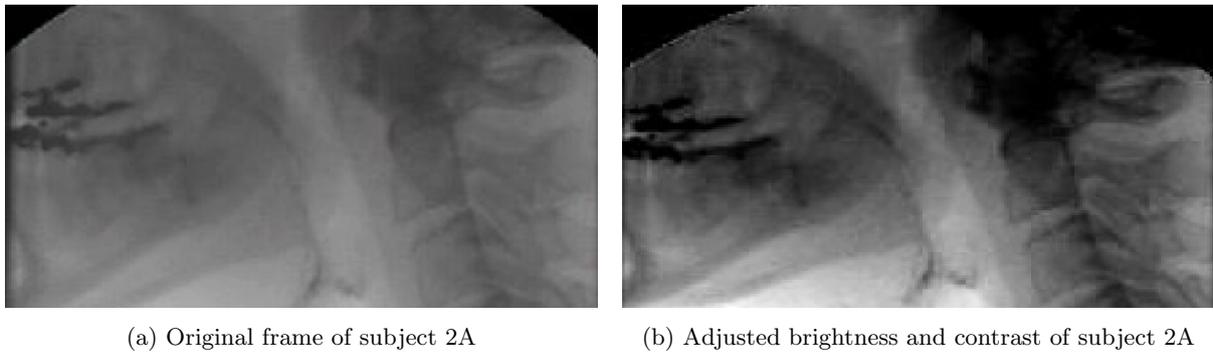


Figure 6.1: Frames from subject 2A showing differences between brightness and contrast.

## 6.2 Non-rigid methods

### 6.2.1 Demons algorithm

Figure 6.2 shows the demon transformation of the moving frame on the basis of the reference frame. This transformed frame shows that the vertebrae are not yet entirely stabilized. Also the demon-transformed frames show deformation in the anatomical structures and in the anatomical swallowing process.

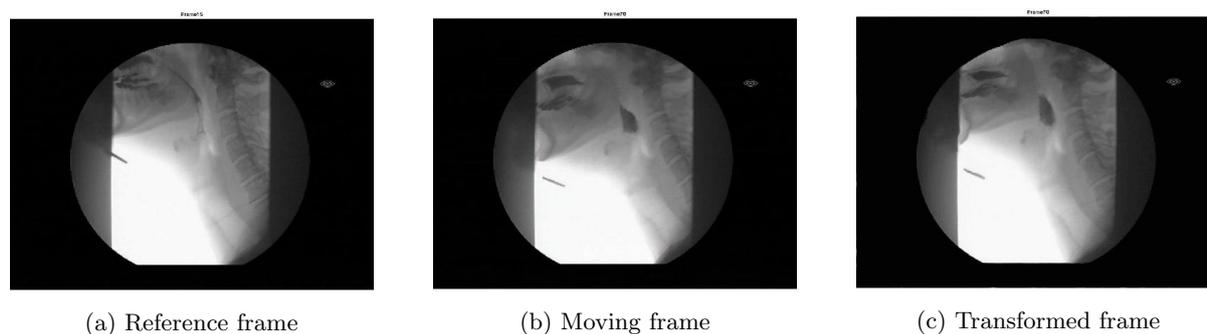


Figure 6.2: The transformation of a frame using the demons algorithm.

### 6.2.2 Mesh-based warping with landmarks

Following, we obtained results using the mesh-based warping algorithm and landmarks. As described in the method we have tried different methods with landmarks, using different amounts and places. The obtained frames were placed on top of each other, the reference frame purple and the edited moving frame green, shown in the tables in Appendix A.<sup>1</sup> This resulted in three different outcomes.

Our first experiment, using a small amount of landmarks ( $< 30$ ), resulted in a video of five frames with the lower vertebrae fairly steadied, but still movement was shown in the jaw, head and upper vertebrae. Also the frames showed a slight deformation, mostly at places with no or a minimal amount of landmarks, such as the mouth and jaw. The tables 6.1- 6.4, in this chapter and the tables in Appendix A show these frames, named in the table as ‘Landmarks  $< 30$ ’. They also show that frames with less than 30 landmarks have not matched all areas entirely, such as the coin placed on the subject’s throat. This indicates that the position of the head within the reference and moving frames differed.

In order to reduce this problems, more landmarks were placed on the frames to make sure that the head was sufficiently covered with landmarks, without covering moving objects (parts of the swallowing process). The result was slightly different from our first result, which can be seen in the enclosed edited fluoroscopy videos in the appendix or the provided videofiles ‘Landmarks  $< 30$ ’ and ‘Landmarks  $> 30$ ’. These show that the head and vertebrae movement was reduced, but the jaw was nevertheless moving significant and deformation was still present. These observations in the videos are also visible in the tables, in the frames named ‘Landmarks  $> 30$ ’. However, it is hard to see, because of the faint colors. Therefore the difference between landmarks  $< 30$  and landmarks  $> 30$  is not very distinct. The greatest difference can be observed in the frames of subject number 1 with consistency A. Especially around the upper vertebra, there is a noticeable green glow in the frames with less than 30 landmarks, which is not observable in the frames with more than 30 landmarks.

At last, landmarks were placed on the moving parts of the swallowing process as well. Meaning, we placed even more landmarks on the head, such as the lower and upper jaw. The result was very different from the other two methods (landmarks  $< 30$  and landmarks  $> 30$ ). This can be seen in the provided videos, named as ‘Landmarks-incl-Jaw’. The head became almost motionless and the swallowing process did not warp in most cases, but video 2A still contained moderate deformation. While the lower jaw appeared steadied, there was still movement in the upper jaw, but this differed per

<sup>1</sup>In some images a green or purple haze is visible in the mouth, oropharynx, hypopharynx and the oesophagus. This shows the bolus at different phases of the swallowing process and the location is therefore different for each frame. This does not refer to movement of the head, so it should not be taken into account during the analysis.

subject. These results are also shown in the tables, named as ‘Landmarks Jaw’. The stabilisation of the lower jaw is best seen in the table, when looking at the coin. The frames that contain additional landmarks on the jaw, show that there is no displacement between the coin of the reference and moving frame anymore. The frames, without landmarks placed on the jaw, point out that there is still a shift between the coins. Furthermore a green glow can be noticed just below the lower jaw in these frames, which is absent in the frames with landmarks on the jaw.

In all three edited videos the bolus was clearly visible, but there was still some head movement between the frames of one video. The edited videos with mesh-based warping using landmarks, who showed the smallest movement of the head were a result of the method using ‘Landmarks Jaw’.

### 6.2.3 Mesh-based warping with segments

After using landmarks for mesh-based warping, we edited the frames by segmenting our regions of interest and creating landmarks. These results are also shown in the tables of the subjects, named as Segmentation. Besides the frames of video 1A and 2A in Tables 6.1, 6.2 and in Tables 6.3, 6.4, we also displayed the frames as a result of videos 1B, 2B, 3A and 4A in Tables A.1, A.2, A.3 and 6.5, in the appendix.

In these frames the movement of the head is minimal, which can be seen in almost all videos such as ‘Subject1A-Segmentation.avi’, ‘Subject1B-Segmentation.avi’, ‘Subject2A-Segmentation.avi’ and ‘Subject4A-Segmentation.avi’. The lower jaw is almost motionless and the upper jaw moves minimally. In most edited video frames, the vertebrae remain at the correct position and there is minimal deformation. However, in the video of subject 2B, movement of the subject’s head can still be noticed and the swallowing process is deformed. Furthermore, the video of subject 3A shows a slight movement of the head as well as minor deformation.

The edited videos are illustrated in the tables. The frames of subject 1A show that the coins are not overlapping and there is a green haze visible around the vertebrae. On the other hand, it looks like the lower jaw is properly stabilised. Looking at the same subject, but at different consistencies (1B), the segmentation method took care of the motion of the lower jaw. The area around the vertebrae is slightly green in some frames within this subject. In Tables 6.3 and 6.4 the frames for subject 2A point out that the vertebrae and jaw are not moving after segmentation. Especially compared to the unedited frames, the movement of the head is eliminated very well.

The frames of subject 2B show a large deformation when playing the video. However, the coin placed on the throat of the subject in the moving frame, did come closer to the coin in the reference frame. Also the Table A.2, in the appendix, shows that there is less of a green and purple glow present in the edited frames, which can be observed when looking at the lower jaw.

The best results of mesh-based warping with segmentation are achieved in the frames of subject 4A. In this video the lower and upper jaw stand extremely still, the coins are overlapping in almost every frame and the vertebrae are less green and purple. Besides this, there is a small deformation.

#### 6.2.4 The absolute values of the difference of consecutive frames

After editing our ten frames using more segmentation regions, we applied the ‘imabsdiff’ method on these frames. The video ‘Subject4A-AbsSegmentation10frames.avi’, containing the difference of the absolute values of the consecutive frames, shows a highlighted moving bolus. Further highlighted areas show that not all frames are matched perfectly. These and other areas are highlighted to a bigger extent in the video ‘Subject4A-AbsOriginal.avi’, the video containing the difference of the absolute values of the consecutive frames of the original video of subject 4A. The video of subject 4A edited with nine segmented areas and ten frames has a minimal amount of head movement and the vertebrae are practically standing still. Also the upper and lower jaw, the coin on the throat and the back of the head look fixed. Despite this, small deformations of the head are still visible, though the amount of deformation seems diminished compared to the video of subject 4A with less segmented areas.

Table 6.1: Transformed frames of subject 1A, frames 1 vs. 2 and 1 vs. 3

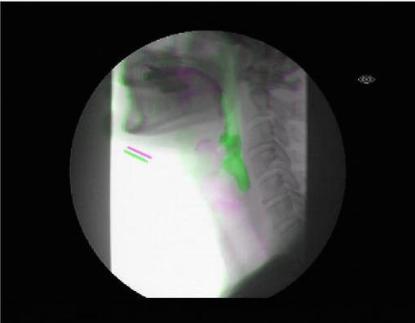
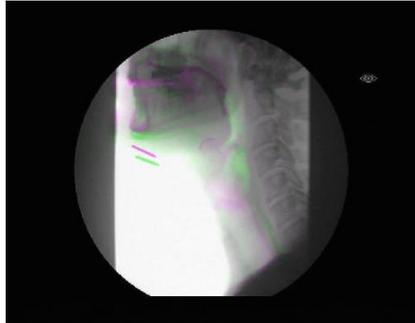
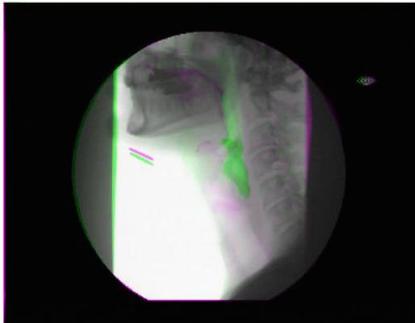
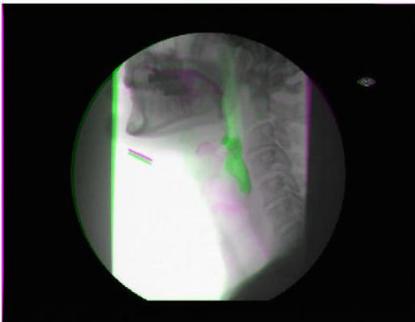
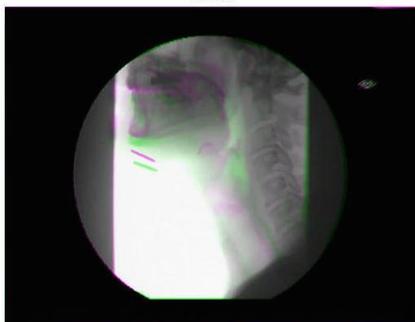
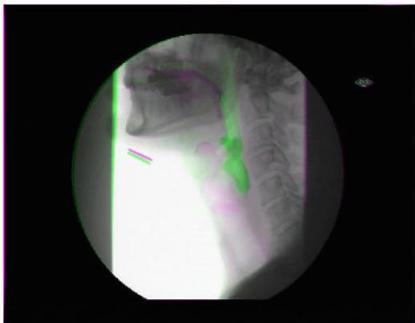
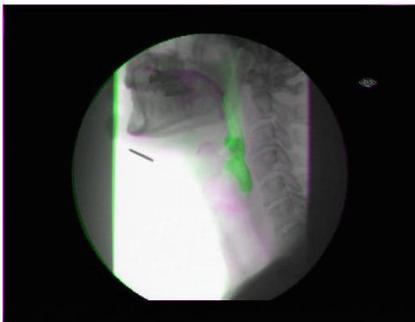
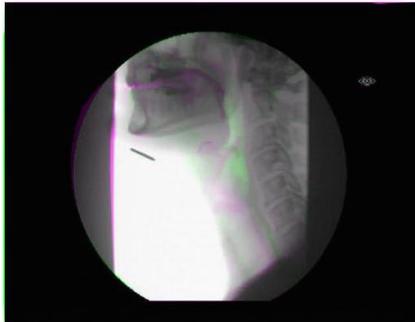
Subject 1A	1 vs. 2	1 vs. 3
Unedited	 <p>Frame16</p>	 <p>Frame16</p>
Segmentation	 <p>Frame16</p>	 <p>Frame16</p>
Landmarks < 30	 <p>Frame16</p>	 <p>Frame16</p>
Landmarks > 30	 <p>Frame16</p>	 <p>Frame16</p>
Landmarks jaw	 <p>Frame16</p>	 <p>Frame16</p>

Table 6.2: Transformed frames of subject 1A, frames 1 vs. 4 and 1 vs. 5

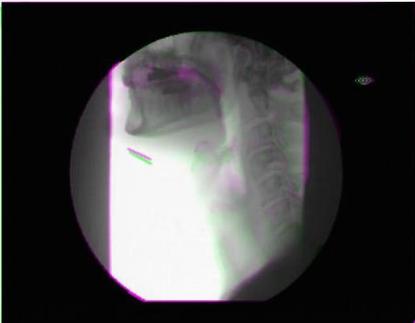
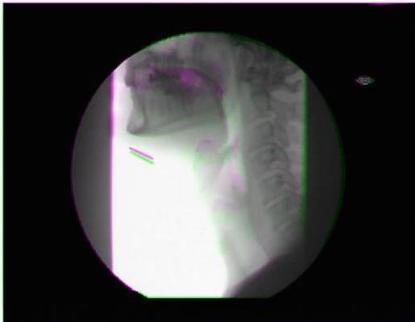
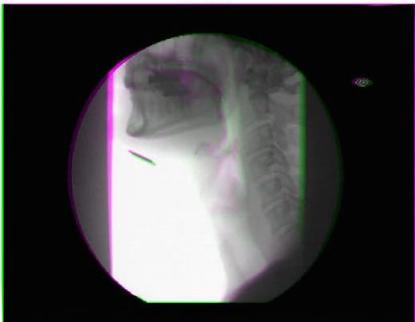
Subject 1A	1 vs. 4	1 vs. 5
Unedited	 <p>Frame15</p>	 <p>Frame15</p>
Segmentation	 <p>Frame15</p>	 <p>Frame15</p>
Landmarks < 30	 <p>Frame15</p>	 <p>Frame15</p>
Landmarks > 30	 <p>Frame15</p>	 <p>Frame15</p>
Landmarks jaw	 <p>Frame15</p>	 <p>Frame15</p>

Table 6.3: Transformed frames of subject 2A, frames 1 vs. 2 and 1 vs. 3

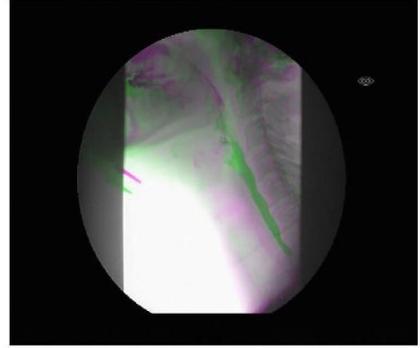
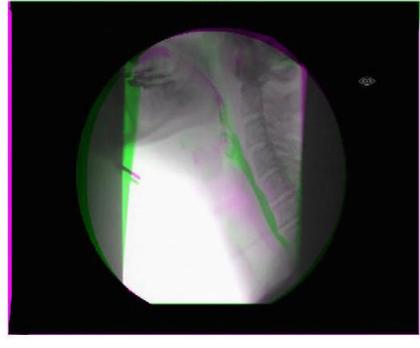
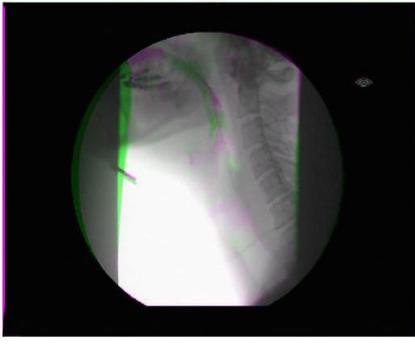
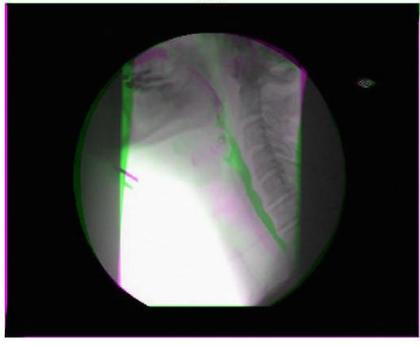
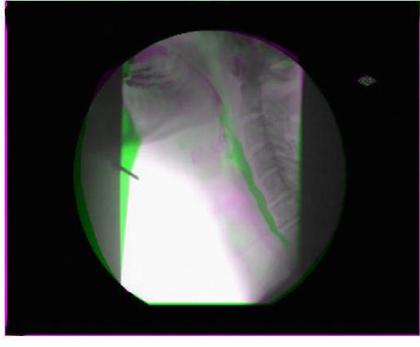
Subject 2A	1 vs. 2	1 vs. 3
Unedited	 <p>Frame 1 vs. 2 comparison, unedited. Shows a circular view of a medical scan with green and magenta overlays. The image is slightly blurred and has some artifacts.</p>	 <p>Frame 1 vs. 3 comparison, unedited. Shows a circular view of a medical scan with green and magenta overlays. The image is slightly blurred and has some artifacts.</p>
Segmentation	 <p>Frame 1 vs. 2 comparison, segmentation. Shows a circular view of a medical scan with green and magenta overlays. The image is sharper and has less noise than the unedited version.</p>	 <p>Frame 1 vs. 3 comparison, segmentation. Shows a circular view of a medical scan with green and magenta overlays. The image is sharper and has less noise than the unedited version.</p>
Landmarks < 30	 <p>Frame 1 vs. 2 comparison, landmarks &lt; 30. Shows a circular view of a medical scan with green and magenta overlays. The image is sharper and has less noise than the unedited version.</p>	 <p>Frame 1 vs. 3 comparison, landmarks &lt; 30. Shows a circular view of a medical scan with green and magenta overlays. The image is sharper and has less noise than the unedited version.</p>
Landmarks > 30	 <p>Frame 1 vs. 2 comparison, landmarks &gt; 30. Shows a circular view of a medical scan with green and magenta overlays. The image is sharper and has less noise than the unedited version.</p>	 <p>Frame 1 vs. 3 comparison, landmarks &gt; 30. Shows a circular view of a medical scan with green and magenta overlays. The image is sharper and has less noise than the unedited version.</p>
Landmarks jaw	 <p>Frame 1 vs. 2 comparison, landmarks jaw. Shows a circular view of a medical scan with green and magenta overlays. The image is sharper and has less noise than the unedited version.</p>	 <p>Frame 1 vs. 3 comparison, landmarks jaw. Shows a circular view of a medical scan with green and magenta overlays. The image is sharper and has less noise than the unedited version.</p>

Table 6.4: Transformed frames of subject 2A, frames 1 vs. 4 and 1 vs. 5

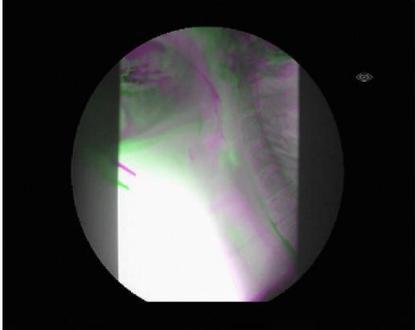
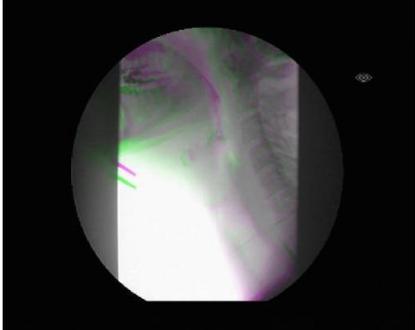
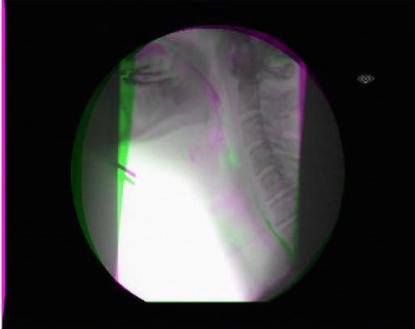
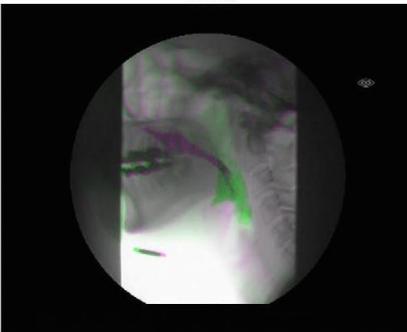
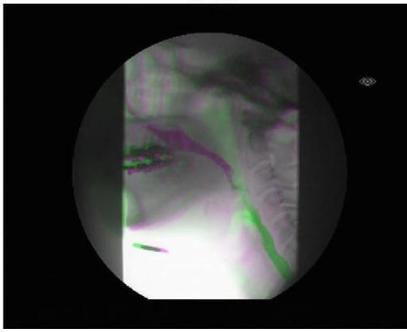
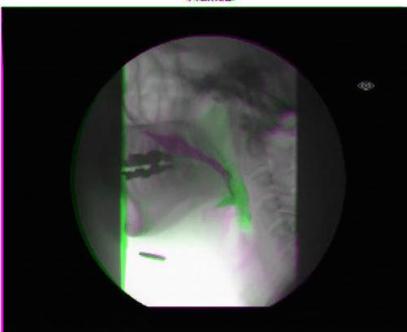
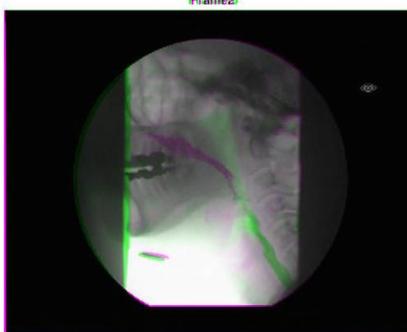
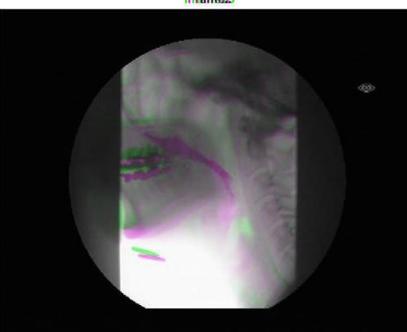
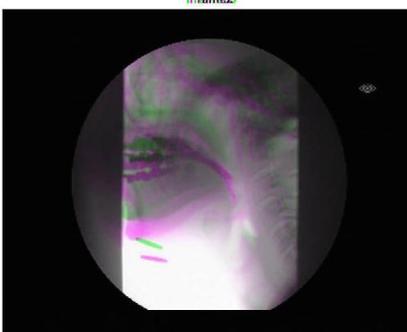
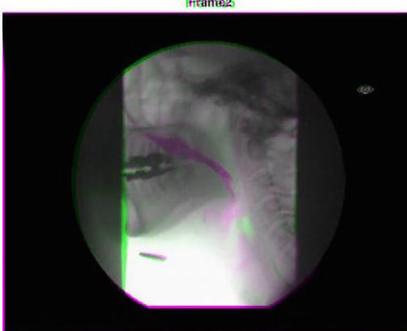
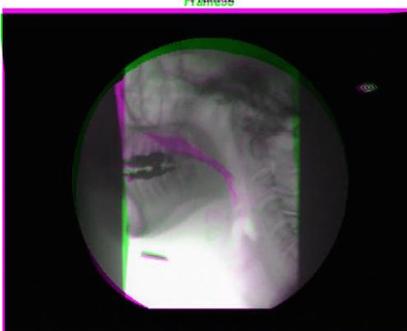
Subject 2A	1 vs. 4	1 vs. 5
Unedited		
Segmentation		
Landmarks < 30		
Landmarks > 30		
Landmarks jaw		

Table 6.5: Transformed frames of subject 4A, frames 1 vs. 2, 1 vs. 3, 1 vs. 4 and 1 vs. 5

Subject 4A	1 vs. 2	1 vs. 3
Unedited	 <p>Frame 2</p>	 <p>Frame 3</p>
Segmentation	 <p>Frame 2</p>	 <p>Frame 3</p>
Subject 4A	1 vs. 4	1 vs. 5
Unedited	 <p>Frame 4</p>	 <p>Frame 5</p>
Segmentation	 <p>Frame 4</p>	 <p>Frame 5</p>



# Discussion

---

IN this chapter, we will review our results and discuss aspects of improvement. As displayed in the results, we managed to eliminate the global head movement in some videos: video 1A, video 1B, video 2A and video 4A. Starting our research, we had hoped to nullify the movement in all fragments. Nevertheless, there were some results that were not satisfactory: video 2B and video 3A. Therefore, we will discuss what could have been improved to achieve better results.

## 7.1 Fragment and frame selection

We selected the fragments of one video, by grading how much a subject moved while swallowing one of the quantities and consistencies. This resulted in Table 5.1. Looking back at our selection procedure, we realised that this method was rather subjective. The fragments we used were those containing 5cc thin and 5cc thick, but had we used other fragments, our results could have been different. Looking at our analysis of different ways people swallow, some patients have difficulty with thin consistencies and some with solid food. We only analysed thin and thick liquids, therefore the results can only be applied to these consistencies.

In our analysis we used five frames from each fragment, which were evenly spread within one fragment and showed different phases of swallowing. This leads to another point of discussion, i.e. there is much time between the frames. In our opinion, this led to some of the displacements we saw in our results. We think that these distortions would be less noticeable, if the time between two frames would be less, which we tried to validate in the video of ten frames. To attain a more accurate and more smoothed result, all circa seventy frames of each fragment should be edited. In order to achieve this, completely manual segmentation should be avoided. If we had found a more efficient method for segmentation, such as semi-automatic segmentation, we may have been able to edit more frames and assemble more data.

## 7.2 Landmarks

The results of the transformation of the frames were not satisfactory after using ‘cpselect’. We tried to place more landmarks on the anatomical structures in order to improve our results. However, in some cases this was not possible due to the fact that not enough anatomical structures were available. It also turned out to be difficult to place an equal amount of landmarks on the frames of every subject, as one subject would have more suitable anatomical structures for landmarks than another. Besides, it was also difficult to manually select the exact same anatomical structure in each frame. As the video continued, sometimes anatomical structures disappeared due to the movement of the patient. Therefore, these structures were not usable. Altogether, manually selecting landmarks is not recommended.

## 7.3 Non-rigid registration methods in Matlab<sup>®</sup>

### 7.3.1 Demons algorithm

We have tried to use the demons algorithm as a transformation model, but it turned out that it did not give the results we wanted. Due to the fact that the demons algorithm does not take anatomical structures into account, the oropharyngeal region is deformed, which makes this method impracticable in our research. However, if in some way, landmarks could be placed on stable structures such as bone, deformation might be prevented.

### 7.3.2 Mesh-based warping

We thought that all radial basis functions were based on splines. Since mesh-based warping uses bicubic splines, we assumed that it was a radial basis function. However, bicubic splines are not per se part of radial basis functions. Therefore our assumption that mesh-based warping is a radial basis function turned out to be false.

It seemed that more artefacts presented itself in the results of subjects that moved excessively. Video 2B showed the most deformation. A possible explanation is the greater movement of the subject's head in comparison to the other subjects. In this research, only videos of healthy subjects were analysed. The difference between the videos of healthy subjects and patients, is that patients tend to show more movement, because they have more difficulty with swallowing. This makes it questionable whether our current transformation model is capable to deal with videos obtained from patients, who are for example treated for head and neck cancer. Therefore, we cannot predict the results of patient's videos and further research should be done to clarify this.

## 7.4 Rigid registration methods in Matlab<sup>®</sup>

After obtaining our results, we saw that the videos that contained the most head movement showed the worst results after editing. Therefore, using a rigid method before the non-rigid registration could have been an improvement. Unfortunately we did not have enough time to examine this.

## 7.5 Segmentation

'ImageSegmenter' requires manual selection of regions of interest. By this means, precise selection was more difficult and did not always generate an impeccable selected segment. In advance, we expected a better result using segmentation compared to manual selection of landmarks. However, we did not see a significant improvement with segmentation relative to placing landmarks on the vertebrae and jaw. This can be explained due to the different structures that we segmented. While segmenting, we mostly segmented the vertebrae, the orbit and the upper and lower jaw. In contrast to the landmarks, which were placed on the coin on the throat and on the back of the head as well. Hereby, the landmarks are more spread amongst the entire frame compared to the created segments.

As previously described in the method, we decided that it would contribute to our research to validate whether more segmented areas would give better results. We analysed the video of subject 4A, in which we selected ten frames and segmented more areas. This resulted in a more fixed head compared

to less segmented areas, but it still showed some deformation. This could be due to more segmented areas, creating more fixed areas and deforming the surrounding areas. Therefore, we think further research is necessary to determine the right amount and place of segmented areas.

## 7.6 ‘ut\_contourfft’

The created segments could not be implemented in our model directly, because every data set has to have the same starting point on the boundary. ‘ut\_contourfft’ should have corrected different starting points, but after running the script and comparing between two frames, we found out that some starting points did not match completely. After resegmentation, sometimes even multiple times, all starting points per segment did match. We did not find out why ‘ut\_contourfft’ made mistakes in matching starting points, but were thus able to correct this by repeating segmentation. A possible explanation could be the adaptation of our initial segments in ‘ImageSegmenter’: evolving the segments with a number of iterations [58].

## 7.7 Validation of the obtained results

After finishing the practical aspects of our research, we found out that answering our second subquestion was not possible with our collected results, because this was not the focus of our research. More data, by editing more frames, could have enlarged the accuracy of our outcome and would have enabled us to validate our edited videos. A small improvement was already visible in the edited video using ten frames: ‘Subject4A-Segmentation10frames.avi’. A complete edited videofragment would have allowed a speech-language therapist to compare the edited fragment to the original fragment. Herewith, we could have received an opinion of the extent to which our research has led to an improvement of the assessment of VFS.

A salient outcome of our research was that subject 4A provided us with fine results, because all major segmentation areas, such as jaw bones and multiple vertebrae, were somewhat zoomed in on and clearly visible in the frame. The settings were different in comparison with other subjects, where larger parts of the trachea and oesophagus were displayed, but the jawbones and other segmentation areas were less focussed on. Therefore, useful regions for segmentation, like the orbit and teeth, were displayed more explicit. These extra landmarks were also spread more uniformly across the frame, what contributes to nullification of the global movement of the head.



# Conclusion

---

AT the beginning of this research, the following research question has been determined:

*To what extent can non-rigid image registration be used to nullify the global movement of healthy subjects during the swallowing process and eventually result in a more objective assessment of the videofluoroscopy?*

In order to be able to answer the main question, two subquestions were set up as well.

## 8.1 Non-rigid image registration method

Our first subquestion was: *Which non-rigid image registration method is best to nullify the global movement of the healthy subject's head?* Looking at the non-rigid methods we compared in our research, mesh-based warping, based on bicubic spline interpolation, turned out to be the best non-rigid method to nullify the global head movement. In contrast to the demons algorithm, mesh-based warping allocates the anatomical structures, such as the vertebrae and takes them into account when applying the transformation. Because the demon algorithm does not consider those structures, deformation of the swallowing process occurs. This outcome is partially in accordance with our hypothesis. The geometric approach turned out to be fitting, but the physical continuum model was not applicable. We expected to use a radial basis function, but eventually we used mesh-based warping instead.

## 8.2 Elimination of the global head movement

Our main goal was to nullify the movement of the subject's head using non-rigid image registration, in order to eventually result in a more objective assessment of the videofluoroscopy recordings. Based on our results, which were acquired using the mesh-based warping transformation model with landmarks, we concluded that using landmarks on anatomical structures, leads to a fixation of the head movement. Subsequently, using more than 30 landmarks, which are also placed on moving structures such as the jaw, leads to a more fixed head movement and less deformation.

The videos, edited with the segmentation method, had varying results. Our results look better when landmarks, acquired with segmentation, are placed on multiple structures spread out across the whole frame. A satisfactory result is for example video 'Subject4A-Segmentation.avi' in which the global movement of the head was removed entirely. An example of a video in which our goal was not achieved, is video 'Subject2B-Segmentation.avi'. In the video of subject 3A, deformation of one of the vertebrae was clearly visible, but the swallowing process was not deformed. Therefore, it will not have a major effect on the assessment of the swallowing process of the subject. Altogether, the segmentation method creates more landmarks than manual selection of landmarks. An adequate amount of landmarks on moving anatomical structures and stable structures, enhanced the extent to which the global movement was frozen.

In the results it is clearly visible that the most fixated head is produced by manual selection of landmarks on the jaw. Due to the fact that the segmentation method indirectly creates a great amount of landmarks, this method is more accurate. We conclude, validated with our video of ten frames, that the best method to nullify the head movement is the mesh-based warping method using segmentation. The segmented areas must correspond with the same regions of interest as the placed landmarks: the vertebrae, the upper and lower jaw, the back of the head, the orbit and the coin on the throat must be segmented and uniformly spread across the whole frame. In this way, the best transformation and therefore the best nullification of the movement of the subject's head will be accomplished.

### 8.3 Accuracy and reproducibility of the VFS assessment

The second subquestion was: *Does the correction of the global movement lead to a more accurate and therefore more reproducible assessment of the videofluoroscopy?* After doing our research, it turned out that our results cannot provide a definite answer to the second subquestion. We cannot guarantee that the nullification of the global head movement will lead to a more accurate and reproducible assessment. To establish this, it would require further research in which for example the edited videos will be used to measure the amount of contrast medium in the different regions of the swallowing process, as described in the problem definition. Since the swallowing process itself is not affected and global head movement is eliminated, we are confident that the aim of measuring the amount of contrast in the trachea and oesophagus is within reach. Further research could enable us to say for sure that the correction of the global movement will lead to a more accurate and therefore more reproducible assessment of the videofluoroscopy recordings.

# Recommendations

---

IN our discussion, we acknowledge that some improvements could be made. In this section we want to elaborate on these improvements and make a few recommendations concerning our research.

First of all, we recommend further research with videos of patients. Taking into account that our method is only tested on videos of healthy subjects, we cannot guarantee our method will be suitable as a first step of ultimately making the assessment of VFS automatic. As it is our belief that patients, who have more trouble with swallowing than healthy subjects, will show more undesired head movements. We did not compare thick and thin consistencies intrasubjectively, because this was not part of our goal to nullify the head movement of the patient. Therefore, we recommend to compare different consistencies in videos of dysphagic patients in further research.

Some videos of our subjects showed more anatomical structures, like vertebrae, and therefore contained more regions we could segment. The amount of visible vertebrae and jawbones differed intersubjectively and intrasubjectively. Therefore, we have a few recommendations for clinical use. First of all, the patient should move as little as possible while swallowing. Furthermore, to obtain the best results using non-rigid image registration, the VFS videofragment should at least contain the following anatomical structures: four vertebrae, the jaw and the orbit. Those structures are needed to produce a satisfying outcome, because our validation video showed that placing more segmented regions results in a better fixation of the head. Bear in mind, the recommended settings should not affect the assessment of the VFS negatively, because important information about the contrast medium in the trachea or oesophagus cannot be lost.

Aside from the above, we have a few recommendations for further research. First of all, we recommend the transformation of all frames instead of only five. Given our limited time, we were not able to edit each frame of every video. We recommend to use segmentation and mesh-based warping in every frame, accordingly making the results more clearly defined. Therefore, as already mentioned in the background, it would be preferable to use automatic segmentation. This would be less time-consuming and errors, due to manual segmentation, could be avoided. In order to validate the clinical relevance of the edited frames, the edited and unedited videos need to be assessed and compared by a SLP. For example, the elevation of the hyoid bone is an important movement of the swallowing function. Therefore it is essential to analyse whether the editing did not deform this important anatomical structure.

Literature research has shown that radial basis functions would probably be suited as a transformation model to nullify the global movement of the subject's head. Therefore, we recommend to examine other radial basis functions, such as thin plate splines, for the transformation of the video fragments, because it might provide better results than the mesh-based warping algorithm [29, 32, 33, 61].

In the future the elimination of the global head movement could be used to determine which percentage of the bolus remains in the oral cavity or oropharynx. Better prediction of functional loss for patients with head and neck cancer would be possible.



# Acknowledgements

---

First of all, we would like to thank our supervisors for their time and effort in guiding us and providing us with feedback, especially F. van der Heijden and L. van der Molen. We would also like to thank the medical staff in the Antoni van Leeuwenhoek Hospital in Amsterdam, for their help, feedback and time to show us the clinical practice. Finally, we want to thank our friends and family who have supported us during our research.



# References

- [1] RNH Datawarehouse. populatieopbouw slikproblemen, 2013.
- [2] Med-Info-Groep. Dysfagie, 2011.
- [3] Hanneke Kalf. Dysfagie, 2013.
- [4] A. Kreeft. *Functional inoperability of oral and oropharyngeal cancer, Chapter 1*. Thesis, 2013.
- [5] P. Garcia-Peris, L. Paron, C. Velasco, C. de la Cuerda, M. Cambor, I. Breton, H. Herencia, J. Verdaguer, C. Navarro, and P. Clave. Long-term prevalence of oropharyngeal dysphagia in head and neck cancer patients: Impact on quality of life. *Clinical Nutrition*, 26(6):710–717, 2007.
- [6] Antoni van Leeuwenhoek. Hoofd-halskanker, 2015.
- [7] Association of Comprehensive Cancer Centers National Cancer Recording. Cijfers over kanker, 2011. (IKNL).
- [8] Intergraal Kankercentrum Nederland. Meest voorkomende soorten kanker, 2015.
- [9] Amy Reinstein. The four phases of the normal adult swallow process, 2010.
- [10] Raj K. Goyal and Hiroshi Mashimo. Physiology of oral, pharyngeal, and esophageal motility, 2006.
- [11] Koichiro Matsuo and Jeffrey B. Palmer. Anatomy and physiology of feeding and swallowing normal and abnormal. *Physical medicine and rehabilitation clinics of North America*, 19(4):691–707, 2008.
- [12] Normal swallow function, 2003.
- [13] Jeri A. Logemann. *Evaluation and Treatment of Swallowing Disorders*. pro-ed, Austin, Texas, 1998.
- [14] Jeri A. Logemann. Behavioral management for oropharyngeal dysphagia. *Folia Phoniatr Logop*, 51(4-5):199212, 1999.
- [15] Lisette van der Molen, Maya A van Rossum, Lori M Burkhead, Ludi E Smeele, and Frans J M. Hilgers. Functional outcomes and rehabilitation strategies in patients treated with chemoradiotherapy for advanced head and neck cancer: a systematic review. *European Archives of Oto-Rhino-Laryngology*, 266(6):889–900, 2009.
- [16] E. C. Ward and C. J. van As-Brooks. *Head and Neck Cancer, Treatment, Rehabilitation, and Outcomes*. Plural Publishing, San Diego, 2007.
- [17] Radiological Society of North America. Video fluoroscopic swallowing exam (vfse), 2013. (RSNA).
- [18] Lisette van der Molen. Introductiegesprek in het antoni van leeuwenhoek met medisch begeleider lisette van der molen, 2015.
- [19] Vinidh Paleri, Justin W. G. Roe, Primo Strojan, June Corry, Vincent Grgoire, Marc Hamoir, Avraham Eisbruch, William M. Mendenhall, Carl E. Silver, Alessandra Rinaldo, Robert P. Takes, and Alfio Ferlito. Strategies to reduce long-term postchemoradiation dysphagia in patients with head and neck cancer: An evidence-based review. *Head & Neck*, 36(3):431–443, 2014.

- [20] Lisette van der Molen, MayaA. van Rossum, LoriM. Burkhead, LudiE. Smeele, CoenR.N. Rasch, and FransJ.M. Hilgers. A randomized preventive rehabilitation trial in advanced head and neck cancer patients treated with chemoradiotherapy: Feasibility, compliance, and short-term effects. *Dysphagia*, 26(2):155–170, 2011.
- [21] Gustavo Nader Marta, Valter Silva, Heloisa de Andrade Carvalho, Fernando Freire de Arruda, Samir Abdallah Hanna, Rafael Gadia, Joo Luis Fernandes da Silva, Sebastio Francisco Miranda Correa, Carlos Eduardo Cintra Vita Abreu, and Rachel Riera. Intensity-modulated radiation therapy for head and neck cancer: Systematic review and meta-analysis. *Radiotherapy and Oncology*, 110(1):9 – 15, 2014.
- [22] D. Rueckert and P. Aljabar. Nonrigid registration of medical images: Theory, methods, and applications [applications corner]. *Signal Processing Magazine, IEEE*, 27(4):113–119, 2010.
- [23] Mark A. Haidekker. *Chapter 11: Image Registration*, book section 11, pages 350–385. Hoboken, NJ : John Wiley & Sons, 2011.
- [24] WR Crum, T Hartkens, and DLG Hill. Non-rigid image registration: theory and practice. *The British Journal of Radiology*, 77(suppl2):S140 – –S153, 2004.
- [25] Vijay Luthra. interpolation, 2015.
- [26] Linda G Shapiro and George C Stockman. *Computer Vision: chapter 10*. Pearson Education (Us), 1 edition, 2001.
- [27] Neeraj Sharma and Lalit M. Aggarwal. Automated medical image segmentation techniques. *J Med Phys*, 35(1):3–14, 2010.
- [28] Mark J. L. Orr. Radial basis function networks, 1996.
- [29] M. Fornefett, K. Rohr, and H. S. Stiehl. Radial basis functions with compact support for elastic registration of medical images. *Image and Vision Computing*, 19(12):87–96, 2001.
- [30] James Ramsay and B. W. Silverman. *Functional Data Analysis*. Springer Series in Statistics. Springer-Verlag, New York, 2 edition, 2005.
- [31] A. Ardeshir Goshtasby. *Image Registration*. Advances in Computer Vision and Pattern Recognition. Springer, London, 2012.
- [32] Fred L Bookstein. Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(6):567–585, 1989.
- [33] K Rohr. Spline-based elastic image registration. *Proc. Appl. Math. Mech.*, 3:3639, 2003.
- [34] Mark Holden. A review of geometric transformations for nonrigid body registration. *IEEE Trans. Med. Imaging*, 27(1):111–128, 2008.
- [35] Igor Yanovsky. *Unbiased Nonlinear Image Registration*. Thesis, 2008.
- [36] William M Wells, Alan Colchester, Scott Delp, Yongmei Wang, and Lawrence H Staib. *Elastic model based non-rigid registration incorporating statistical shape information*, volume 1496 of *Lecture Notes in Computer Science*, pages 1162–1173. Springer Berlin Heidelberg, 1998.
- [37] Themis P. Exarchos, Athanasios Papadopoulos, and Dimitrios I. Fotiadis. *Handbook of Research on Advanced Techniques in Diagnostic Imaging and Biomedical Applications*. IGI Global, Hershey, PA, USA, 2009.

- 
- [38] Roger Claypoole, Jim Lewis, Srikrishna Bhashyam, and Kevin Kelly. Elec 539, image morphing, 1997.
- [39] Pramod Potluri, Krishna Sagiraju, and Venkatachalem Tubati. Image morphing: Feature based, view and mesh, 2004.
- [40] Mark Steyvers. Morphing techniques for manipulating face images. *Behavior Research Methods, Instruments, & Computers*, 31(2):359–369, 1999. Behavior Research Methods, Instruments, & Computers 0743-3808.
- [41] George Wolberg. Image morphing: a survey. *The Visual Computer*, 14(8-9):360–372, 1998. The Visual Computer 0178-2789.
- [42] Sky Mckinley and Megan Levine. Cubic spline interpolation.
- [43] A.M. Kreeft. *Functional inoperability of oral and oropharyngeal cancer, Chapter 4*. Thesis, 2013.
- [44] Karen H. O’Neil, Mary Purdy, Janice Falk, and Lanelle Gallo. The dysphagia outcome and severity scale. *Dysphagia*, 14(3):139–145, 1999.
- [45] Bonnie Martin-Harris, MartinB Brodsky, Yvonne Michel, DonaldO Castell, Melanie Schleicher, John Sandidge, Rebekah Maxwell, and Julie Blair. Mbs measurement tool for swallow impairmentmbsimp: Establishing a standard. *Dysphagia*, 23(4):392–405, 2008.
- [46] Anne Marijn Kreeft, Lisette van der Molen, Frans Hilgers, and Alfons J Balm. Speech and swallowing after surgical treatment of advanced oral and oropharyngeal carcinoma: a systematic review of the literature. *European Archives of Oto-Rhino-Laryngology*, 266(11):1687–1698, 2009.
- [47] L Kuil. *2D Motion Analysis of Swallowing in Videofluoroscopy*. Thesis, 2013.
- [48] ChattanoogaGroup. Dysphagia fact sheet, 2005.
- [49] Margaret Walshe. Oropharyngeal dysphagia in neurodegenerative disease. *Journal of Gastroenterology and Hepatology Research*, 3(10):1265–1271, 2014.
- [50] Richard D. Wilson and Evan C. Howe. A cost-effectiveness analysis of screening methods for dysphagia after stroke. *PM&R*, 4(4):273–282, 2012.
- [51] Mission statement of the virtual therapy consortium.
- [52] MATLAB<sup>®</sup>. *version 8.5.0.197613 (R2015a)*. The MathWorks Inc., Natick, Massachusetts, 2015.
- [53] MATLAB<sup>®</sup>. *Modifying Images,version 8.5.0.197613 (R2015a)*. The MathWorks Inc., Natick, Massachusetts, 2015.
- [54] MATLAB<sup>®</sup>. *Control Point Registration,version 8.5.0.197613 (R2015a)*. The MathWorks Inc., Natick, Massachusetts, 2015.
- [55] MATLAB<sup>®</sup>. *Automatic registration, version 8.5.0.197613 (R2015a)*. The Mathworks Inc., Natick, Massachusetts, 2015.
- [56] Naive. Mesh-based image warping, 2014.
- [57] MATLAB<sup>®</sup>. *ROI-Based Processing, version 8.5.0.197613*. The MathWorks Inc., Natick, Massachusetts, 2015.
- [58] MATLAB<sup>®</sup>. *Image Segmentation, version 8.5.0.197613 (R2015a)*. The MathWorks Inc., Natick, Massachusetts, 2015.

- [59] F. Van der Heijden. Function `ut_contourfft`, 2006.
- [60] MATLAB<sup>®</sup>. *GPU Computing, version 8.5.0.197613*. The Mathworks Inc., 2015 2015.
- [61] L. Zagorchev and A. Goshtasby. A comparative study of transformation functions for nonrigid image registration. *Image Processing, IEEE Transactions on*, 15(3):529–538, March 2006.

## APPENDIX A

# Results

---

THE results we obtained by editing the videofluoroscopy videos are examined in the discussion and conclusion. The edited videos are classified and named as followed:

- Manual selection of landmarks
  - Landmarks-less-30:
    - \* Subject1A-less30.avi
    - \* Subject2A-less30.avi
  - Landmarks-more-30
    - \* Subject1A-more30.avi
    - \* Subject2A-more30.avi
  - Landmarks-incl-Jaw
    - \* Subject1A-Jaw.avi
    - \* Subject2A-Jaw.avi
- Segmentation
  - Subject1A-Segmentation.avi
  - Subject1B-Segmentation.avi
  - Subject2A-Segmentation.avi
  - Subject2B-Segmentation.avi
  - Subject3A-Segmentation.avi
  - Subject4A-Segmentation.avi
  - Validation
    - \* Subject4A-Segmentation10frames.avi
    - \* Subject4A-AbsSegmentation10frames.avi
    - \* Subject4A-AbsOriginal.avi

These edited video files can be obtained by contacting the group members of this research. The remainder of the results are shown in the tables in this appendix.

Table A.1: Transformed frames of subject 1B, frames 1 vs. 2, 1 vs. 3, 1 vs. 4, 1 vs. 5

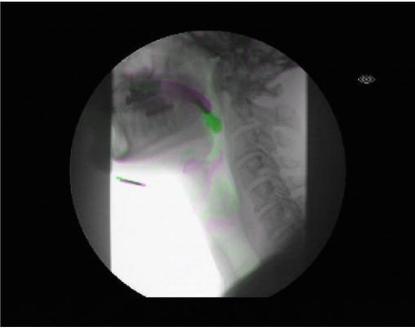
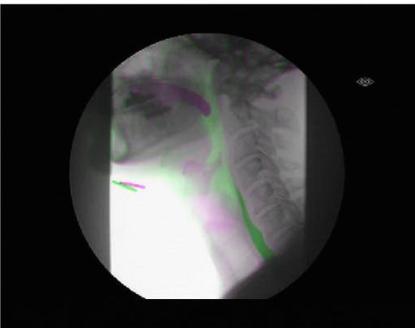
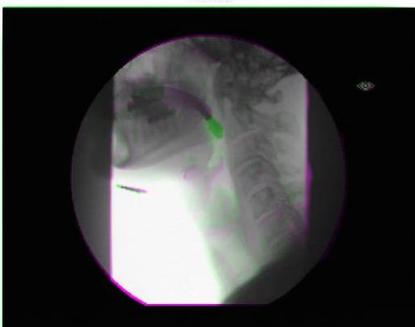
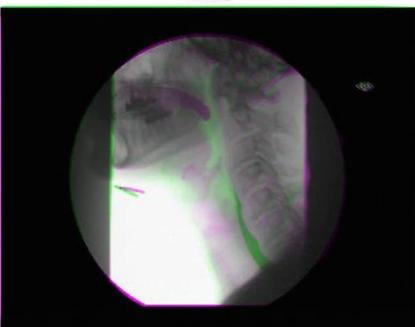
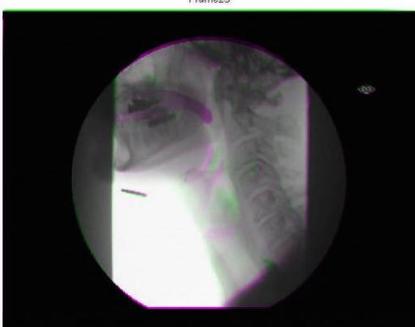
Subject 1B	1 vs. 2	1 vs. 3
Unedited	 <p>Frame25</p>	 <p>Frame25</p>
Segmentation	 <p>Frame25</p>	 <p>Frame25</p>
Subject 1B	1 vs. 4	1 vs. 5
Unedited	 <p>Frame25</p>	 <p>Frame25</p>
Segmentation	 <p>Frame25</p>	 <p>Frame25</p>

Table A.2: Transformed frames of subject 2B, frames 1 vs. 2, 1 vs. 3, 1 vs. 4, 1 vs. 5

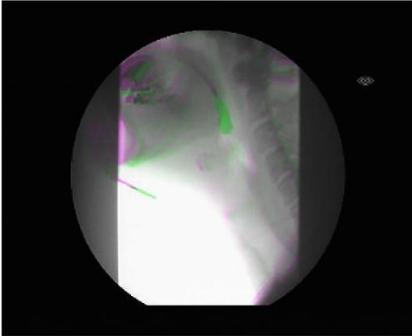
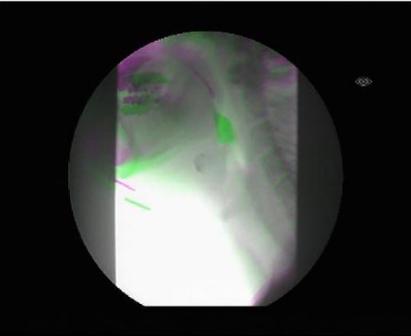
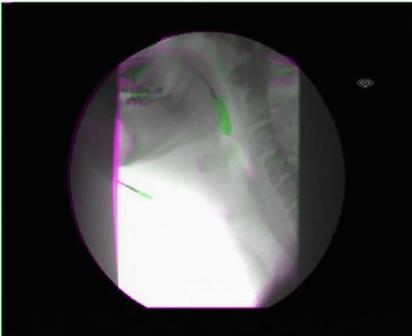
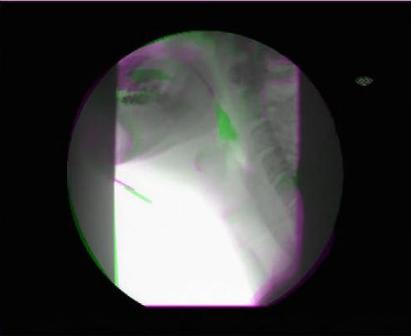
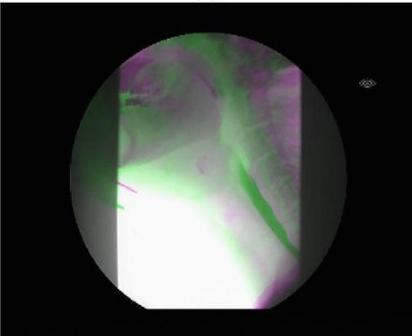
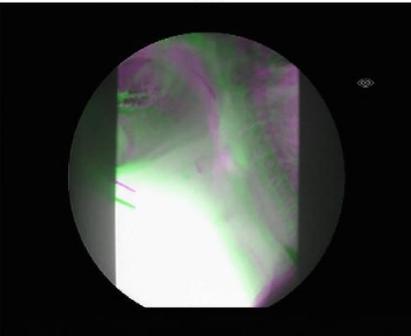
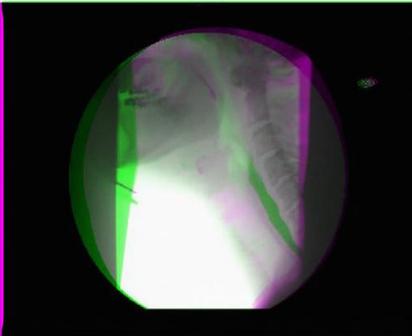
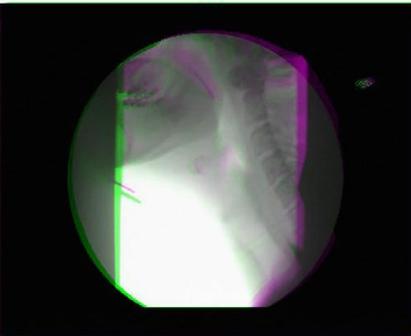
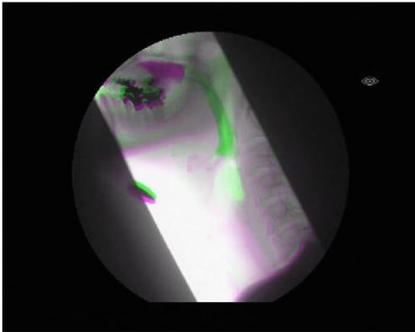
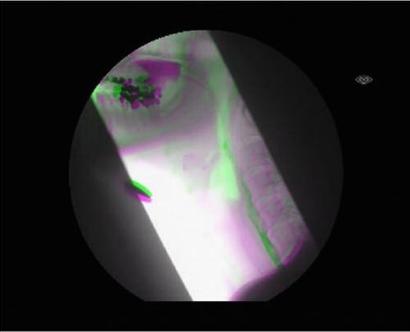
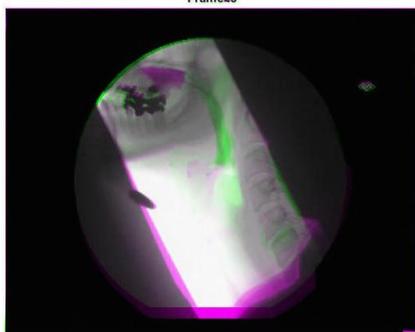
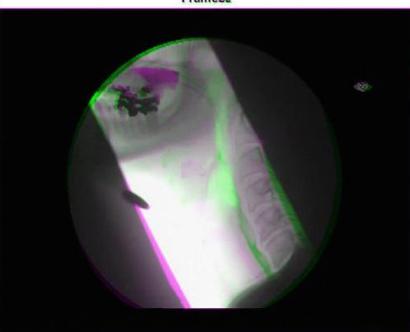
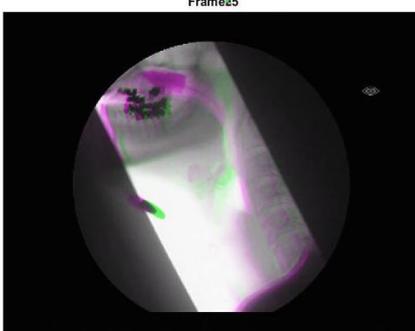
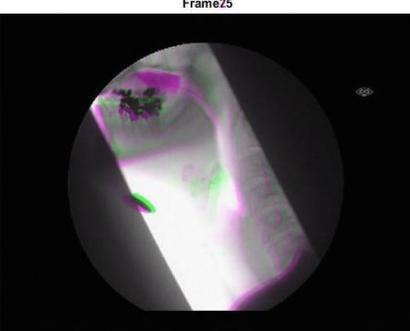
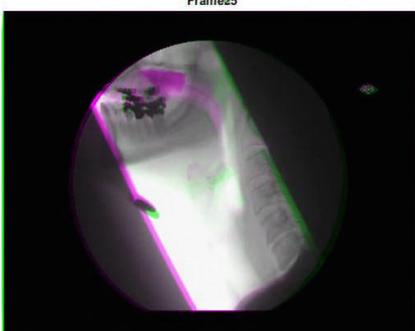
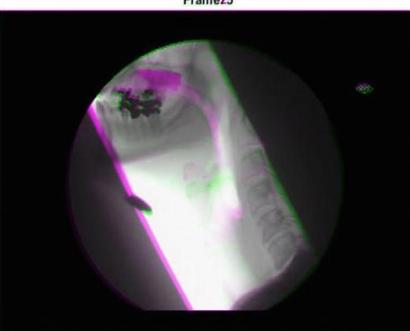
Subject 2B	1 vs. 2	1 vs. 3
Unedited	 <p>Frame 02</p>	 <p>Frame 02</p>
Segmentation	 <p>Frame 02</p>	 <p>Frame 02</p>
Subject 2B	1 vs. 4	1 vs. 5
Unedited	 <p>Frame 02</p>	 <p>Frame 02</p>
Segmentation	 <p>Frame 02</p>	 <p>Frame 02</p>

Table A.3: Transformed frames of subject 3A, frames 1 vs. 2, 1 vs. 3, 1 vs. 4, 1 vs. 5

Subject 3A	1 vs. 2	1 vs. 3
Unedited	<p style="text-align: center;">Frame25</p> 	<p style="text-align: center;">Frame25</p> 
Segmentation	<p style="text-align: center;">Frame25</p> 	<p style="text-align: center;">Frame25</p> 
Subject 3A	1 vs. 4	1 vs. 5
Unedited	<p style="text-align: center;">Frame25</p> 	<p style="text-align: center;">Frame25</p> 
Segmentation	<p style="text-align: center;">Frame25</p> 	<p style="text-align: center;">Frame25</p> 

# Mesh-based warping script

The first script defines the function ‘mesh-based warping’ [56].

```

1 function [Q] = mesh_based_warping(sourceImage, targetImage, possource, posstarget,
   alpha, beta)
2 % warp the source image to a target using a mesh-based warping method.
3 % sourceImage: RGB format input source image
4 % targetImage: RGB format input target image which has the same size with source image
5 % num_points: number of feature points
6 % alpha: coefficient
7 %     alpha = 1 -> output = target image
8 %     alpha = 0 -> output = source image
9
10 xxA=possource(1,:);
11 yyA=possource(2,:);
12 xA=xxA(:);
13 yA=yyA(:);
14
15 xxB=posstarget(1,:);
16 yyB=posstarget(2,:);
17 xB=xxB(:);
18 yB=yyB(:);
19
20 A=sourceImage;
21 B=targetImage;
22
23 height = size(sourceImage,1);
24 width = size(sourceImage,2);
25
26 height2 = size(targetImage,1);
27 width2 = size(targetImage,2);
28
29 if(height~=height2 || width~=width2)
30     error('Input images should have same size.')
31 end
32
33 xA = [xA;1; width; width; 1];
34 yA = [yA;1;1; height; height];
35
36 xB = [xB;1; width; width; 1];
37 yB = [yB;1;1; height; height];
38
39 % create a intermediate grid
40 C = zeros(height, width, 3);
41 xC = alpha*xA + (1-alpha)*xB;
42 yC = alpha*yA + (1-alpha)*yB;
43 triC = delaunay(xC,yC);
44 ntri = size(triC,1);
45
46 % allocate memory for x, y coordinators
47 xCA = zeros(height, width);
48 yCA = zeros(height, width);
49 xCB = zeros(height, width);
50 yCB = zeros(height, width);

```

```

51 [X,Y] = meshgrid(1:width,1:height);
52
53 % warp the intermediate grid to source and target grid
54 for k = 1:ntri
55     [w1,w2,w3,r] = inTri(X, Y, xC(triC(k,1)), yC(triC(k,1)), xC(triC(k,2)), yC(triC(k,2)),
56         xC(triC(k,3)), yC(triC(k,3)));
57     w1(~r)=0;
58     w2(~r)=0;
59     w3(~r)=0;
60     xCA = xCA + w1.*xA(triC(k,1)) + w2.*xA(triC(k,2)) + w3.*xA(triC(k,3));
61     yCA = yCA + w1.*yA(triC(k,1)) + w2.*yA(triC(k,2)) + w3.*yA(triC(k,3));
62     xCB = xCB + w1.*xB(triC(k,1)) + w2.*xB(triC(k,2)) + w3.*xB(triC(k,3));
63     yCB = yCB + w1.*yB(triC(k,1)) + w2.*yB(triC(k,2)) + w3.*yB(triC(k,3));
64 end
65 % interpolate each point by using 'interp2' function
66 VCA(:,:,1) = interp2(X,Y,double(A(:,:,1)),xCA,yCA);
67 VCA(:,:,2) = interp2(X,Y,double(A(:,:,2)),xCA,yCA);
68 VCA(:,:,3) = interp2(X,Y,double(A(:,:,3)),xCA,yCA);
69
70 VCB(:,:,1) = interp2(X,Y,double(B(:,:,1)),xCB,yCB);
71 VCB(:,:,2) = interp2(X,Y,double(B(:,:,2)),xCB,yCB);
72 VCB(:,:,3) = interp2(X,Y,double(B(:,:,3)),xCB,yCB);
73
74 % cross-dissolve
75 C = beta*VCA + (1-beta)*VCB;
76
77 % convert double to uint8 format and display
78 %imshow(uint8(C));
79 Q=uint8(C);
80 end
81
82 function [w1,w2,w3,r] = inTri(vx, vy, v0x, v0y, v1x, v1y, v2x, v2y)
83 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
84 % inTri checks whether input points (vx, vy) are in a triangle whose
85 % vertices are (v0x, v0y), (v1x, v1y) and (v2x, v2y) and returns the linear
86 % combination weight, i.e., vx = w1*v0x + w2*v1x + w3*v2x and
87 % vy = w1*v0y + w2*v1y + w3*v2y. If a point is in the triangle, the
88 % corresponding r will be 1 and otherwise 0.
89 %
90 % This function accepts multiple point inputs, e.g., for two points (1,2),
91 % (20,30), vx = (1, 20) and vy = (2, 30). In this case, w1, w2, w3 and r will
92 % be vectors. The function only accepts the vertices of one triangle.
93 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
94     v0x = repmat(v0x, size(vx,1), size(vx,2));
95     v0y = repmat(v0y, size(vx,1), size(vx,2));
96     v1x = repmat(v1x, size(vx,1), size(vx,2));
97     v1y = repmat(v1y, size(vx,1), size(vx,2));
98     v2x = repmat(v2x, size(vx,1), size(vx,2));
99     v2y = repmat(v2y, size(vx,1), size(vx,2));
100     w1 = ((vx-v2x).*(v1y-v2y) - (vy-v2y).*(v1x-v2x))./...
101         ((v0x-v2x).*(v1y-v2y) - (v0y-v2y).*(v1x-v2x))+eps;
102     w2 = ((vx-v2x).*(v0y-v2y) - (vy-v2y).*(v0x-v2x))./...
103         ((v1x-v2x).*(v0y-v2y) - (v1y-v2y).*(v0x-v2x))+eps;
104     w3 = 1 - w1 - w2;
105     r = (w1>=0) & (w2>=0) & (w3>=0) & (w1<=1) & (w2<=1) & (w3<=1);
106 end

```

mesh\_based\_warping.m

The second script uses the previous stated function, mesh-based warping, and is used to obtain the warped frames.

```

1 clear all
2 close all
3
4 fig1=imread('referenceframe1'); % load frames
5 fig2=imread('movingframe2');
6 fig3=imread('movingframe3');
7 fig4=imread('movingframe4');
8 fig5=imread('movingframe5');
9
10 load ('refpoints1')
11 load ('movpoints2')
12 load ('movpoints3')
13 load ('movpoints4')
14 load ('movpoints5')
15
16 alpha=0.99;
17 beta=0.01;
18 warpfig2 = mesh_based_warping(fig1 , fig2 , refpoints1 ', movpoints2 ', alpha , beta);
19 warpfig3 = mesh_based_warping(fig1 , fig3 , refpoints1 ', movpoints3 ', alpha , beta);
20 warpfig4 = mesh_based_warping(fig1 , fig4 , refpoints1 ', movpoints4 ', alpha , beta);
21 warpfig5 = mesh_based_warping(fig1 , fig5 , refpoints1 ', movpoints5 ', alpha , beta);
22 %%
23 outputVideo = VideoWriter('slikkensubject.avi');
24 outputVideo.FrameRate = 3;
25 open(outputVideo)
26 writeVideo(outputVideo , fig1)
27 writeVideo(outputVideo , warpfig2)
28 writeVideo(outputVideo , warpfig3)
29 writeVideo(outputVideo , warpfig4)
30 writeVideo(outputVideo , warpfig5)
31 close(outputVideo)
32 slikken2avi = VideoReader('slikkensubject.avi');
33 ii = 1;
34 while hasFrame(slikken2avi)
35     mov(ii) = im2frame(readFrame(slikken2avi));
36     ii = ii+1;
37 end
38
39 f = figure(3);
40 f.Position = [150 150 slikken2avi.Width slikken2avi.Height];
41
42 ax = gca;
43 ax.Units = 'pixels';
44 ax.Position = [0 0 slikken2avi.Width slikken2avi.Height];
45
46 image(mov(1).cdata , 'Parent' ,ax)
47 axis off
48
49 movie(mov,10 ,slikken2avi .FrameRate)

```

domeshwarp.m



# Demon registration script

---

This script uses the function ‘imregdemons’ to warp the frames.

```

1 close all
2
3 load 'referenceframe1'
4 load 'movingframe2'
5 load 'movingframe3'
6 load 'movingframe4'
7 load 'movingframe5'
8
9 fixed = imread('referenceframe1');
10 moving = imread('movingframe2'); % repeat for the consecutive frames
11
12 fixed = rgb2gray(fixed);
13 moving = rgb2gray(moving);
14
15 figure
16 imshowpair(fixed, moving, 'montage');
17 figure
18 imshowpair(fixed, moving)
19
20 moving = imhistmatch(moving, fixed);
21
22 [~, movingReg] = imregdemons(moving, fixed, [500 400 200], 'AccumulatedFieldSmoothing', 1.3);
23
24 fixedGPU = gpuArray(fixed);
25 movingGPU = gpuArray(moving);
26
27 fixedHist = imhist(fixedGPU);
28 movingGPU = histeq(movingGPU, fixedHist);
29
30 [~, movingReg] = imregdemons(movingGPU, fixedGPU, [500 400 200], 'AccumulatedFieldSmoothing', 1.3);
31
32 movingReg = gather(movingReg);
33
34 figure
35 imshowpair(fixed, movingReg)
36 figure
37 imshowpair(fixed, movingReg, 'montage')
38
39 %%
40 fig1 = imread('referenceframe1');
41 fig2 = imread('movinframe2tranformed');
42 fig3 = imread('movinframe3tranformed');
43 fig4 = imread('movinframe4tranformed');
44 fig5 = imread('movinframe5tranformed');
45
46 outputVideo = VideoWriter('swallowingdemons.avi');
47 outputVideo.FrameRate = 2;
48 open(outputVideo)
49 writeVideo(outputVideo, fig1)
50 writeVideo(outputVideo, fig2)

```

```
51 writeVideo(outputVideo, fig3)
52 writeVideo(outputVideo, fig4)
53 writeVideo(outputVideo, fig5)
54 close(outputVideo)
55 slikkenavi = VideoReader('swallowingdemons.avi');
56 ii = 1;
57 while hasFrame(slikkenavi)
58     mov(ii) = im2frame(readFrame(slikkenavi));
59     ii = ii+1;
60 end
61
62 f = figure(3);
63 f.Position = [150 150 slikkenavi.Width slikkenavi.Height];
64
65 ax = gca;
66 ax.Units = 'pixels';
67 ax.Position = [0 0 slikkenavi.Width slikkenavi.Height];
68
69 image(mov(1).cdata, 'Parent', ax)
70 axis off
71
72 movie(mov, 1, slikkenavi.FrameRate)
```

demons\_algorithm.m