

The background of the cover features abstract, low-poly orange geometric shapes. A large, complex shape on the left side resembles a stylized figure or a cluster of triangles, while a more elongated, chain-like structure is positioned at the top right. These shapes are composed of various shades of orange and white, creating a modern, architectural feel.

MASTER THESIS

# Towards holographic navigation for paediatric nephron-sparing Wilms tumour surgery: a feasibility study

Q.D. Eyck, BSc

<sup>1</sup>TechMed Centre, University of Twente

<sup>2</sup>Department of Paediatric Surgery, Princess Máxima Centre for paediatric oncology

12 March, 2021

## EXAMINATION COMMITTEE

Dr. ir. F. van der Heijden<sup>1</sup> (Chairman and technical supervisor)

Prof. dr. M.H.W.A. Wijnen<sup>2</sup> (Medical supervisor)

Drs. R.J. Haarman<sup>1</sup> (Process supervisor)

M. Fitski, MSc<sup>2</sup> (Technical supervisor)

T. Boers, MSc<sup>1</sup> (External member)

## Preface

In September 2014 ben ik begonnen aan de Bachelor Technische Geneeskunde in Enschede. Ondanks een roerige en zware periode heb ik mijn bachelor binnen de “normale” drie jaar gehaald. Na afronding van mijn eerste jaar van de master Medical Imaging and Interventions had gehaald, vond ik het nog te vroeg om meteen aan mijn stages te beginnen. Om er een leerzame ervaring van te maken heb ik gekozen voor de studiereis naar Japan. Na deze reis ben ik begonnen aan mijn vier korte stages. Tijdens deze stages werd het me duidelijk dat ik graag zou willen afstuderen op de chirurgie, maar in welk ziekenhuis of in welke stad wist ik nog niet. Uiteindelijk kwam ik de opdracht van prof. Wijnen tegen. Het feit dat het een opdracht op de kinderchirurgie bij het Princes Máxima Centrum betrof sprak mij zeer aan. Daarnaast was ik zeer te spreken over de spirit die in het Princes Máxima Centrum hangt. Alles is erop gefocust om ervoor te zorgen dat de zorg voor kinderen met kanker te verbeteren met als uiteindelijk doel dat er geen enkel kind meer hoeft te overlijden aan kanker. Dit doel wordt nagestreefd door veel onderzoek te doen en door innovatief te zijn. Het afstudeerproject waarop ik had gereageerd was nog heel breed geformuleerd. Tijdens het kennismakingsgesprek bleek dat ik eigenlijk kon kiezen tussen drie projecten. Daarvan was er één gericht op het gebruik van Augmented Reality (AR) bij kinderen met nierkanker. Ik had totaal geen ervaring met AR, maar dit heeft me er niet van weerhouden om na het kennismakingsgesprek deze uitdaging aan te gaan.

Tijdens het afgelopen jaar heb ik me vol ingezet om een AR applicatie te ontwikkelen zodat dit uiteindelijk een grotere rol kan gaan spelen in de behandeling van kinderen met kanker. Ik heb veel overleg gehad met het 3DLab van het Radboudumc in Nijmegen. Dankzij hun hulp ben ik in staat geweest om deze applicatie te ontwikkelen tot het systeem wat het nu is. Naast de ontwikkeling van de AR applicatie heb ik mij ook klinisch verder kunnen ontwikkelen in het Prinses Máxima Centrum. Ondanks dat dit jaar anders verliep door het Covid-19 virus, heb ik toch veel klinische ervaring op kunnen doen. Door mijn kennis over oncologische kinderchirurgie te verbreden heb ik kunnen bijdragen aan het zorgproces op de operatiekamer en in het preoperatieve proces. Van tevoren had ik niet kunnen bedenken dat ik daadwerkelijk chirurgische handelingen zou uitvoeren op kinderen.

Dit eindresultaat was niet bereikt zonder de hulp van mijn begeleiders. Graag wil ik daarom Marc Wijnen bedanken voor de mogelijkheid om in dit centrum mijn afstudeerstage te lopen en om mij te kunnen ontwikkelen op de afdeling kinderchirurgie. Daarnaast Matthijs Fitski voor de dagelijkse begeleiding en de sparmomenten. Uiteraard wil ik ook Ferdi van der Heijden bedanken voor de technische en wetenschappelijke input en gesprekken. Rian, Eva en Marijn bedankt voor de leuke en vooral inzicht gevende procesbijeenkomsten.

Ik ben tot het inzicht gekomen dat ik mij in de toekomst graag wil inzetten om de gezondheidszorg te verbeteren middels het implementeren van nieuwe technieken.

Ik wens u veel leesplezier.

Quinten Eyck

Utrecht, maart 2021

## Abstract

Wilms Tumour (WT) affects one out of 10,000 children before the age of fifteen. Nephron-sparing surgery (NSS) can be performed under specific rules, due to the risk of positive resection margins. This necessitates additional radiotherapy and possibly additional chemotherapy. Therefore, surgeons remain hesitant to perform NSS.

We have developed and evaluated an Augmented Reality (AR) surgical navigation application in order to gain more intraoperative insight in the patient-specific anatomy. 3D meshes created from the preoperative MRI scan are imported into the HoloLens 2 (HL2). Multiple critical structures can be divided in these 3D meshes: kidney, tumour and vessels. The accuracy of this application is analysed based on the tracking ability of QR-coded markers, and the ability of the user to apply the point-based registration method based on a 3D printed kidney phantom. Subsequently, a clinical validation study is performed to analyse the user-friendliness and usability in a more clinical setting.

During the accuracy measurement of the QR-coded marker recognition three different dimensions of QR-markers are used. The marker with a dimension of 5x5 cm appeared to have a superior accuracy compared to smaller dimensions. The application after the point-based registration has a mean accuracy of 3.16 millimetre, based on 50 marked points. The NASA raw task load index (RTLX), performed in the clinical validation study, showed results compared to other medical tasks. The surgeons, who performed the clinical validation, are enthusiastic about the possibilities of the HoloLens during nephron-sparing surgery. The registration method is usable during NSS.

To conclude, we have developed an AR surgical navigation application based on a point-based registration method and a QR-marker tracking method for intraoperative use to guide surgeons during NSS of WT patient. We hope to further develop and study this technique to eventually implement this technique into the clinical workflow during NSS.

## Contents

Preface.....	i
Abstract .....	ii
List of figures .....	v
List of tables .....	vi
List of abbreviations .....	vii
1 Introduction.....	1
1.2 Research aim .....	2
1.3 Thesis design .....	2
2 Clinical Background .....	3
2.1 Wilms tumour.....	3
2.2 Treatment protocol .....	3
2.3 Medical imaging .....	4
2.4 Surgical procedure.....	5
3 Technical background.....	7
3.1 Current visualisation of medical imaging .....	7
3.2 Augmented reality.....	7
3.3 Optical see-through head-mounted displays.....	8
3.3.1 Microsoft HoloLens 2 .....	8
3.4 Medical applications of the HoloLens .....	9
4 Development of the AR surgical navigation application .....	10
4.1 Overview.....	10
4.2 Design options .....	12
4.2.1 Registration methods .....	12
4.2.2 Tracking methods .....	12
4.3 The preoperative stage .....	13
4.4 Patient registration.....	13
4.5 Tracking .....	15
4.6 Realization .....	15
4.7 Clinical workflow .....	16
5 Part 1: Accuracy of QR-coded marker recognition.....	18
5.1 Introduction.....	18
5.2 Materials and methods .....	18
5.3 Results .....	20

5.4 Discussion .....	23
5.5 Conclusion .....	24
6 Part 2: Accuracy based on a 3D-printed kidney phantom.....	25
6.1 Introduction.....	25
6.2 Materials and methods .....	25
6.3 Results .....	27
6.4 Discussion .....	29
6.5 Conclusion .....	30
7 Part 3: Clinical validation .....	31
7.1 Introduction.....	31
7.2 Materials and methods .....	31
7.3 Results .....	33
7.4 Discussion .....	35
7.5 Conclusion .....	36
8 Overall discussion .....	37
9 Overall recommendations / future perspectives .....	39
9.1 Short-term technical improvements .....	39
9.2 Long-term technical improvements .....	39
9.3 Next steps to clinical implementation.....	40
10 Overall conclusion .....	41
11 References.....	42
12 Appendix.....	49
12.1 UMBRELLA staging criteria .....	49
12.2 Questionnaire clinical validation .....	51
12.3 Abstract CARS 2021 .....	55

## List of figures

Figure 1	Visualisation of the HoloLens 1 and the HoloLens 2, derived from Microsoft. [56], [59]	9
Figure 2	A schematic overview of the proposed workflow of the developed surgical navigation application for intraoperative use to guide surgeons during NSS of WT patients.	11
Figure 3	A schematic overview of the various used coordinate systems and the transformations between them.	11
Figure 4	A visualisation of the main menu of the home-build HoloLens 2 application.	16
Figure 5	A visualisation of the patient scene of the home-build HoloLens 2 application.	17
Figure 6	This graph visualised the mean measured distance of the different QR-coded marker dimensions per position and the standard deviation per movement.	22
Figure 7	This figure shows the visualisation of the holographic overlay seen through the HoloLens 2. This overlay consists of the kidney (salmon), tumour (cyan), artery (red), and a positional marker (purple) and is visualised on top of the 3D printed model. Additionally, the QR-marker used for tracking and the QR-marker placed onto the surgical pointer (blue) are visible. The surgical pointer is used to pinpoint the anatomical landmark for the HoloLens 2. There is a slight offset of the overlay due to a difference in height of the camera of the HoloLens 2 and the eyes of the user.	26
Figure 8	This graph visualised the distance error in millimetre of all observers per marked point.	27
Figure 9	This graph consist of boxplots of all errors of all points combined per observer. This figure shows the clinically maximally allowed error (dashed line), and the variation per observer. Three points are significantly different of the other points set by the observer, however, these points all are within the clinically maximally allowed error.	28
Figure 10	This image consists of two figures, in the left figure the phantom is visualised ready for use. The right part of the image showed one of the surgeons during the registration process.	33
Figure 11	Visualisation after applying the registration method onto the abdominal phantom, visualised through the HoloLens 2.	33
Figure 12	This graph visualised the RTLX-scores per observer per category of the Nasa Raw TLX based on the use of the home-build augmented reality application.	34

## List of tables

Table 1	This table visualised the mean distances between two consecutive positions of the QR-marker, with dimensions of 5x5 cm, in centimetres and compared with the real distance of 4.74 centimetre.	20
Table 2	This table visualised the mean distances between two consecutive positions of the QR-marker, with dimensions of 4.5x4.5 cm, in centimetres and compared with the real distance of 4.74 centimetre.	21
Table 3	This table visualised the mean distances between two consecutive positions of the QR-marker, with dimensions of 4x4 cm, in centimetres and compared with the real distance of 4.74 centimetre.	21
Table 4	In this table the mean values of the measured angle of the different QR-markers, measured in degrees, are visualised.	22
Table 5	Results of the mean error of the point per observer, with their standard deviation.	28
Table 6	This table visualised the scores of the Likert-scale questions given by the five paediatric surgeons after using the home-build augmented reality application.	34

## List of abbreviations

WT	Wilms tumour
bWT	Bilateral Wilms tumour
uWT	Unilateral Wilms tumour
SIOP-RTSG	The Renal Tumour Study Group of the International Society of Paediatric Oncology
COG	Children's Oncology Group
NSS	nephron-sparing surgery
AR	Augmented Reality
RN	radical nephrectomy
2D	Two-dimensional
3D	Three-dimensional
VR	Virtual Reality
WAGR	Wilms tumour Aniridia Genitourinary anomalies and mental Retardation
DDS	Denys-Drash syndrome
BWS	Beckwith-Wiedemann syndrome
US	Ultrasound
MRI	Magnetic Resonance Imaging
HMD	Head-mounted display
HL	HoloLens
OR	Operating room
QR	Quick response
EM	Electromagnetic
RMSE	Root-mean-square error
RTLX	Raw NASA task load index
IR	Infrared



## 1 Introduction

One out of 10,000 children are affected by a Wilms tumour (WT) before the age of fifteen. [1], [2] This corresponds to over 90% of all malignant paediatric kidney neoplasms and 5% of all paediatric cancers. [3] WT is usually diagnosed between the age of one and five years. The mean age of presentation of a Wilms tumour is between three and four years. [1],[2] WT can occur in one kidney (unilateral) or in both kidneys (bilateral). The latter occurs in 5-8% of all cases. Bilateral Wilms tumours (bWT) lead to a worse overall survival (80%) in comparison to unilateral Wilms Tumours (uWT) (90%). Unfortunately, 15% of all WT patients relapse, often during the first two years after treatment. [4]

In Europe, patients suffering from WT are diagnosed and treated according to the UMBRELLA protocol. This protocol has been developed by The Renal Tumour Study Group of the International Society of Paediatric Oncology (SIOP-RTSG). Patients suffering from WT in America are treated according to the protocol of the Children's Oncology Group (COG). [5], [6] The Umbrella protocol is commonly used in the Princess Máxima Centre.

Patients suffering from bWT are treated with nephron-sparing surgery (NSS), when possible. Chen et al. (2020) [7] suggested that, NSS has multiple advantages over a radical nephrectomy (RN) in some cases. According to their meta-analysis the renal function of patients who are treated with RN is decreased, caused by a reduction in glomerular filtration. Besides a better renal function, patients treated with NSS have a longer survivability as well. [7] The main concern is the risk of positive resection margins. [8]–[10] This results in upstaging of the tumour and an increase in the local recurrence rate. [11] Upstaging the tumour necessitates postoperative chemotherapy and possibly radiotherapy.

Another concern of an increased risk of positive resection margins is the usages of two-dimensional (2D) preoperative medical imaging of complex three-dimensional (3D) structures. There are three underlying causes which increase this concern. First, It acquires a specific skill to translate these 2D images into a mental 3D reconstruction by the surgeons. Secondly, this mental 3D reconstruction is necessary to prepare and perform the surgical procedure. The surgeon relies on this internal representation and transformation of the 2D images during the planning of surgery. This planning is based on the 2D images and the surgical goal. Lastly, the surgeon relies on his or her ability to recognise anatomical structures often with from unusual angles never seen in anatomy textbooks intraoperatively. Therefore, the performance of the surgical procedure relies on the ability of the surgeon for spatial representations and processes. [12], [13]

Augmented Reality (AR) could be a technical solution to help surgeons visualise anatomy in 3D instead of 2D, and therewith helping the ability of the surgeon spatial understanding. AR is a variation of virtual reality (VR). Regarding VR technologies, users can only observe the artificial world that is created by the VR technique and while using the VR technique they cannot observe the real world. On the contrary, AR technologies are developed to project digital information, such as audio, images or video, onto the real world. This digital information is not noticeable for other people who are not using AR. Therefore, AR can be thought of as a blend of both the real and the virtual world. [14]–[16] AR can be used to visualise 3D representations of MRI footage. In this representation different structures can be differentiated, which can provide added information to decide the surgical path of the NSS. AR applications can be used during the surgical planning or intraoperatively. A sufficient intraoperatively AR system contains both a registration and a tracking method. Registration methods are used to align the special properties of the real world and the virtual world. After the registration method is applied, all data is stored in the same coordinate system. Tracking methods are used for this continuous live position and orientation update.

## 1.2 Research aim

The research objective of this thesis is to develop and evaluate an AR surgical navigation application based on a registration and tracking method for the visualisation of a Wilms tumour during NSS with an accuracy of maximally five millimetre. To achieve this goal, three sub-questions have to be resolved first.

1. What are the best registration and tracking method?
2. What is the tracking and registration capability of this AR surgical navigation application?
3. Are the paediatric oncologic surgeons satisfied with the developed AR surgical navigation application?

The first sub-question will be assed with a literature study. During this literature study different registration and tracking methods used in different medical applications will be assed.

After creating the AR surgical navigation application, it has to been evaluated to investigate the accuracy. The clinically maximally allowed error for this validation was determined to be five mm by the surgical team of the Princess Máxima Centre. The accuracy will be evaluated using this clinically maximally allowed error. The accuracy of the AR application will be evaluated purely based on the technical ability and when used by different observers.

The last sub-question will be evaluated during a clinical validation study. This clinical validation study will be performed by all five paediatric oncology surgeons of the Princess Máxima Centre.

## 1.3 Thesis design

The clinical relevance is further explained in the second chapter of this study. First, an overview of risk factors of Wilms tumours is given. Followed by the UMBRELLA treatment protocol. Subsequently, the medical imaging and surgical procedure will be clarified in depth.

In chapter three the technical background is explained more detailed. First, the current visualisation of the medical imaging is explained, followed by the introduction of augmented reality and the Optical See-Through Head Mounted Displays and finally some examples of medical augmented reality systems are mentioned.

Chapter four gives an insight on the development of the AR surgical navigation system. In this chapter the whole process of this application is stated. Started with the section: “overview”. Subsequently, different registration and tracking methods will be evaluated and choice of the used registration and tracking method will be explained. Followed by an explanation on how these methods technically work and how these are integrated into the AR application. The third part of this chapter states the programs used to create this AR application. The last part of this chapter states the clinical workflow and explains how the application will be used.

The fifth chapter is the first of three measurements chapters. In this chapter the measurements on the minimal dimensions of QR-markers, which still can be detected with the HoloLens 2, are performed. Followed by the second measurement chapter of this thesis, chapter six. The measurements performed in this chapter will determine the accuracy of the AR application. The last measurement chapter is based on the clinical validation. In this chapter all five paediatric surgeons of the Princess Máxima Centre are asked to test the application. These three chapters are all subdivided into an introduction, methods and materials, results, discussion and a conclusion section.

This thesis will be concluded with an overall discussion, future perspectives and conclusion section.

## 2 Clinical Background

### 2.1 Wilms tumour

Wilms tumour is a common paediatric tumour, which is linked to early kidney development and is associated with other kidney abnormalities. [17] Multiple conditions are known risk factors for the development of a Wilms tumour. According to the British National Register of Childhood Tumours, 9% of the children who suffer from WT have an underlying congenital malformation. [18] Another study, Sanpakit et al. (2013) showed that 13% of the long-term survivors are diagnosed with a syndrome, which is the highest percentage of any paediatric cancer. [19] Congenital malformations with the highest risk of a WT are syndromes demonstrated with mutations of the Wilms Tumour 1 gene, such as WAGR (Wilms tumour Aniridia Genitourinary anomalies and mental Retardation) and the Denys-Drash syndrome (DDS). Other risk factors are familial Wilms Tumours and Childhood overgrowth syndromes, such as Beckwith-Wiedemann syndrome (BWS). [20] Due to the high risk of developing WT, children suffering from WAGR, DDS and BWS are routinely monitored with ultrasound to identify WT in an early stage, thus providing a better prognosis. Moreover, these children also have an increased risk for multiple malignancies. Therefore, they are, if possible, treated in a similar approach as children with bWT, to contain as much healthy kidney tissue as possible, which will be explained in section 2.4. [21]

### 2.2 Treatment protocol

When the patients arrive into the hospital, they receive a diagnostic ultrasound and MRI. After confirming the presence of a renal tumour, the UMBRELLA treatment protocol starts. The UMBRELLA protocol distinguished five stages (I, II, III, IV and V) of WT. The staging is described in Appendix 12.1. These stages are divided into three subgroups, based on postoperative histology, namely low, intermediate and high risk. Most WTs show triphasic patterns of blastemal, epithelial and stromal tissue. WTs are classified as low risk when all three histological characteristics are present and there is no diffuse aplasia present. [22] WTs are classified as high risk if aplasia is present, or if it is a blastemal Wilms tumour type. [23]. The stage and risk factor affects the treatment choice of the WT, resulting in three main groups, local disease: metastatic disease and bilateral disease. [4],[23]

Local disease (stage I, II and III) is treated with a preoperative actinomycin D and vincristine for newly diagnosed patients older than 6 months. This treatment downstages the tumour and reduces the need of radiotherapy and doxorubicin by 20%. [24] Children < 6 months old with an unilateral localized kidney tumour receive direct nephrectomy, because of the higher chance of a congenital mesoblastic nephroma or malignant rhabdoid tumour than a Wilms tumour. [23] The patients ≥ 6 months old receive a radical nephrectomy (RN) after the neoadjuvant chemotherapy. The surgical procedure will be explained in section 2.4. The UMBRELLA protocol advocates the sampling of seven locoregional lymph nodes after neoadjuvant chemotherapy and surgery. The purpose of this sampling is to provide an accurate staging of the disease.

17% of WT patients are diagnosed with metastatic disease (stage IV). Wilms tumour metastases are defined as haematogenous metastases to the lung, liver or other places, or positive extra-abdominal lymph nodes. Lung metastases are most frequently observed. These patients are also treated with preoperative actinomycin D, vincristine and additionally doxorubicin. After six weeks, reassessment based on preoperative imaging is performed. With this treatment 61-67% of patients experience a total metastatic response preoperatively. After the preoperative chemotherapy, RN is performed. [23]

The third and last group is patients suffering from bilateral Wilms tumours (stage V). End-stage renal disease is clinically the most relevant morbidity for this group. Several factors are thought to have a negative influence the change of end-stage renal disease. These include under-treatment caused by under-staging, increased incidence of anaplasia and a delay in disease control. These factors resulting in an extended treatment period, consisting of chemotherapy, without effect. [25] A study of Aronson et al. (2011) proved that the best functional renal outcome is gained after nephron-sparing surgery. [26] Before NSS is suitable for patients suffering from BWT, they are treated with vincristine and actinomycin D for a maximum period of 12 weeks. After six weeks an evaluation is performed to see if NSS is feasible already. [27]

The standard of care for patients with stage I-IV is RN. [23] Under certain conditions, however, NSS is also considered for patients suffering from uWT. These conditions are: a small tumour volume (<300mL), tumour located in one kidney pole, no external invasion of the tumour, and the expectation that there will be a substantial remaining kidney function. [7] Suggested is that NSS has multiple advantages over a RN in some cases. According to the meta-analysis of Chen et al. (2020), the renal function of patients who are treated with RN is decreased, caused by a reduction in glomerular filtration. Besides the improved renal function, patients treated with NSS had a longer survivability in this review as well. [7] Another disadvantage is the 2-3% change of the development of a metachronous contralateral Wilms tumour after RN. [2], [28], [29]

The treatment of uWT with NSS is controversial due to multiple concerns. The main concern is the risk of positive resection margins. [8]–[10] This results in upstaging of the tumour and an increase in the local recurrence rate and a decrease in survival prognosis. [11] Upstaging of the tumour necessitates postoperative chemotherapy and possibly radiotherapy. These consequences will harm the surrounding tissue and the kidney. Besides the physical harm, positive tumour margins are a mental burden as well. Patients and parents experience more stress due to the diagnosis and the uncertainty of the future perspectives. Another concern is that studies investigating NSS for patients with an uWT are subject to significant selection bias. The patient populations between NSS and RN are likely to be different. Patients who underwent NSS generally had a smaller tumour than patients in the RN cohort, which corresponded with a smaller incidence of complications. [7], [30], [31] Therefore Cost et al. (2012) performed a histological study to investigate the percentage of patients who could receive NSS. They retrospectively reviewed all uWT patients treated with a RN after preoperative chemotherapy during the period between 1985 and 2010 in Dallas. They reviewed the pathological specimens of these patients. 78 children were included in this study, of which 19 (24.4%) fitted all predefined criteria for NSS. However not all predefined criteria are suitable as a preoperative criteria. [32]

### 2.3 Medical imaging

According to the UMBRELLA protocol, patients suspected of an abdominal mass receive an ultrasound (US) procedure as a first line modality. US can differentiate between solid or cystic masses and besides the location of origin of the mass can be determined. During the US procedure the contralateral kidney and the liver are carefully examined as well. If possible, the inferior vena cava and renal vein are examined for intravenous tumour thrombus. Subsequently, a diagnostic Magnetic Resonance Imaging (MRI) scan is made. [33], [34]

After the neoadjuvant chemotherapy, preoperative medical imaging is used to determine the surgical treatment and to decrease the risks of complications. An example of these preoperative techniques is MRI. The MRI is used for the assessment of the tumour, and the effectiveness of the neoadjuvant therapy, and to determine the surgical plan. Additionally, a Computer Tomography scan is made, for

monitoring pulmonary nodulus. The lung parenchyma is the most common location for metastases for Wilms tumours. [35], [36] As stated previously, NSS in uWT is only performed if the tumour volume is lower than 300 mL, and if the surgeon expects to preserve substantial amount of healthy kidney tissue. The MRI is used to determine the volume of the tumour. However, the assessment of feasibility of NSS requires extensive anatomical knowledge of the parenchymal and tumour vasculature, tumour location, the relation between tumour and surrounding tissue. [23]

The current MRI protocol in the Princess Máxima Centre is based on the 1.5 Tesla (Achieva, Philips Medical Systems, Best, Netherlands) and consists of coronal T2 weighted sequence, a fat-suppressed T1 weighted sequence, a diffusion weighted imaging and at last a fat-suppressed T2 weighted sequence. Before the contrast agent (Gadovist, Bayer Pharma, Berlin, Germany, at a dose of 0.1 mmol/kg body weight) is administered, a non-contrast magnetic resonance angiogram is made. Subsequently, a 4D contrast enhanced MRA is performed. Depending on the level of cooperation of the patients, they are awake, sedated or under anaesthesia. [37]

During a NSS procedure, US is used for understanding of the tumour infiltration in the renal parenchyma. The US probe is protected with a sterile cover, which allows the surgeon to use the US probe directly onto the kidney and tumour. The intraoperative ultrasound is used to identify the tumour boundaries, healthy parenchyma, vasculature and the urine collection system. However, the occurrence of positive margins is approximately one third for NSS. [38] This is mainly attributed to the difficult delineation between positive margin and artifacts caused by the resection. This limits the use of intraoperative US as a live imaging technique during tumour resection. [9], [38]–[40]

The preoperative MRI and perioperative US have one major disadvantage in common. These imaging modalities visualize the region of interest on a two-dimensional (2D) monitor. The disadvantage of this 2D visualisation is the lack of ability to visualise the tumour depth perception. [41] A deficient tumour depth perception complicate the determination of the tumour location and filtration depth by the surgeon. When the surgeon determines this location wrongly, this can results in a positive tumour margin.

In conclusion, it remains unclear when and how NSS can be performed safely in patients with uWT. Surgeons are dependent on their own 3D interpretation of 2D preoperative imaging. This increases the hesitation to perform a NSS on patients suffering from uWT. Research on improving the preoperative visualisation techniques is required to decrease this hesitation.

### 2.4 Surgical procedure

After neoadjuvant chemotherapy, patients require local surgical control of the tumour, either through RN or NSS. In both cases, a long transverse supraumbilical incision is made, which permits an overview of the whole abdomen and if necessary, the contralateral kidney, which reduce the chance of rupture and tumour spillage. During a RN the kidney has to be mobilized from the surrounding tissue. After mobilization of additional critical structures such as the renal artery, renal vein and ureter, the renal vein and artery are ligated. This ligation can be performed for maximally 30 minutes, after 30 minutes ischemia will occur. [42] There is no evidence that advocates a specific order. The ureter has to be ligated as low as possible, to increase the chance that all migrated tumour cells in the ureter are removed from the patient. All three structures are clipped with titanium clips to mark the resection areas for possible postoperative radiotherapy. After the resection of the kidney, lymph nodes are sampled. [33], [43]

During nephron-sparing surgery, the kidney is mobilized as well. The perirenal fat is kept intact and removed with the tumour, to reduce the risk of tumour rupture. The kidney wall is surrounded with either, sponges, pads or a plastic bowel bag to prevent seeding in case of tumour spillage. The tumour is excised completely with a margin of normal renal tissue. The dissection plane is decided using intraoperative ultrasound. However, intraoperative ultrasound can only be used before the actual tumour resection. Once the resection is performed, it is difficult to delineate the tumour margin from the artifacts caused by the diathermy. [9] The surgeon has to interpret the ultrasound footage for the localisation of the resection margin. This combined with the maximal resection time of 30 minutes, results in time constraints and self-dependency of the surgeon. Partly because of this, as stated previously approximately 30% of the resection margins are positive after a NSS. [9], [38]–[40] The tumour is staged as a stage III, after positive resection margins. This results into the consequences stated in section 2.2.

To prevent extreme blood loss the renal artery can be clamped. However, the obstruction of blood flow should be kept to a minimum to reduce ischemia of the healthy renal parenchyma. When the tumour is resected, the haemostasis can be achieved using absorbable sutures, which closes the small vessels. Again, the resection lymph nodes are subsequently sampled to accurately stage the tumour. [10], [44]

## 3 Technical background

### 3.1 Current visualisation of medical imaging

Surgical plans are based on 2D imaging techniques which requires the surgeon to translate these images into the three-dimensional (3D) anatomy of the patient. Various methods have been proposed to perform this translation using 3D visualisation techniques of 2D preoperative images. [12], [41], [45] These 3D visualisations help to improve the understanding of the location of the tumour and critical anatomical structures. One method for 3D visualisation of 2D medical images is a 3D printed model of the kidney, tumour, and critical anatomical structures. [41], [46] This model is increasingly used prior to surgery. The anatomical structures are extracted from the preoperative MRI scan. Using these models in kidney surgery has proven to improve surgical outcome by reducing blood loss, lowering the risk of intraoperative complications. The effect of these 3D models in paediatric kidney surgery specifically, is not yet known due to the lower incidence of paediatric surgery. [37], [41]

The extraction of the anatomical structures of the preoperative MRI is done using a segmentation technique. Optimal segmentation of different anatomical structures requires dedicated MRI sequences. In the Princess Máxima Centre the following MRI sequences are used; a postcontrast fat-suppressed T1 weighted sequence is used for the segmentation of the parenchymal tissue of the kidney and the tumour. Arteries are segmented based on a non-contrast MRA sequence. Parenchymal arteries and veins are differentiated, using a T2 weighted MRI as an overlay. All segmentations are performed in 3DSlicer (The Slicer Community, version 4.11.20200930, 2020, United States of America), an open-source software package. 3DSlicer consists of multiple segmentation tools, which are used in different methods for the different structures. The grow-cut algorithm is used for the kidney and tumour. Threshold based segmentation is used of the arteries, and the veins and urine collection system are segmented manually. [37]

Nonetheless, the surgeon is still required to interpret the structures of the 3D printed model and translate this knowledge to the surgical field. Additionally, the infiltration depth of the tumour cannot be visualised using 3D printed models. These disadvantages could possibly be solved with the use of Augmented Reality (AR). [47]

### 3.2 Augmented reality

Augmented reality is already integrated in our lives. An example of AR used on daily basis is the use of the heads-up display in airplanes. The HUD displays information onto the cockpit window. This contains information on the magnitude and speed of the plane, and it shows an artificial horizon. During the WWII the British military used this technique to display the radar information onto the windshield, making it one of the oldest applications of AR. [15], [48] A second example is the use of AR on mobile phones. Multiple AR applications exist, for instance the Pokémon Go app, or apps developed to visualise new furniture in your room. [48] The Pokémon Go app uses the real geographical space visualising Pokémon in the real world. Moreover, players are required to physically cover distances to interact with the Pokémon's and other players. [49]

While AR is already widely used, still there are some application dependant challenges of which object recognition and sensor accuracy are the most prominent. The alignment of the virtual world and the real world is often not perfect, due to the absence of a registration methods. AR devices need multiple sensors to work properly, for instance, cameras or other visual sensors, GPS, accelerometers, and wireless sensors. These sensors vary in precision and accuracy and are dependent on sufficient lighting, which makes indoor use of AR especially challenging. [15], [50]



### 3.3 Optical see-through head-mounted displays

As stated above, AR can be used for real-time interaction and 3D registration of virtual and real objects [14], [16]. Methods for the display of AR footage vary even within specific disciplines. Regarding surgery, the mostly used method is to superimpose the preoperative images on live captured camera imagery, and to display the resulting on a monitor. Another display method is through the so-called head-mounted display (HMD). HMDs are glasses equipped with special projectors, head tracking cameras, and depth cameras to display the 3D models on the glass. [51]

The use of HMDs has benefits compared to other AR devices. The ergonomics are improved, as the virtual content is showed right in front of the user, obviating the need to shift the visual focus between the object of interest and an external monitor. By improving ergonomics, the focus of the users improves, which is extremely important during highly focussed task such as surgery. Another benefit of HMDs is that these devices are connectable to Wi-Fi, which makes it accessible to communicate with other users or to facilitate other information. [52] Due to these benefits, HMDs are the most accepted technique for AR display in medical research [53].

Multiple HMDs are available, for example the Google Glass and the Microsoft HoloLens (HL). The Google Glass was one of the first HMDs entering the market. This device consists of a pair of glasses equipped with a small screen and a camera. The medical field was one of the first fields to adopt the Google Glass in 2013, using it for educational, telecommunicational and documentational purposes, e.g. livestreaming a surgical procedure to a remote party. [52], [54] A drawback of the Google Glass is that it does not track his environment, making it unsuitable for visualisation of 3D structures. [55]

Microsoft released the counterpart of the Google Glass in 2016, the HoloLens. The HoloLens has a robust design with a large screen for both eyes. [56] The display runs on the Windows 10 computer system. Different from the Google Glass, the HoloLens has the ability to track his environment. Another advantage of the HL is the ability to have an unhindered view, a so called optical see-through head-mounted display. This is especially an advantage for clinical use, because this provides the surgeon with an unhindered view of the surgical scene. Based on a systematic review, Qian at al. (2017) suggest that the HL is the best optical see-through head-mounted display that is currently available to use in the OR [57].

Limitations of HMDs include battery power and patient privacy. The battery power is not sufficient for an entire surgical procedure. The battery life of the Google Glass facilitate streaming for only 30-40 minutes, the HoloLens 2 for 2-3 hours. [52], [58], [59] Another limitation is the privacy of patient information when used as a medical application. HMDs are connected to the internet, which makes it prone to hacking. [52], [58]

#### 3.3.1 Microsoft HoloLens 2

Microsoft released the HoloLens first generation (HL1) on 30 March 2016. It was world's first commercially available untethered 3D augmented reality device. However, the HL1 could be improved both technically and ergonomically. [56], [60]

by the end of 2019, the second generation of the HL was released (HL2). This second generation implemented improvement in different areas compared to the HL1. The HL2 improved ergonomically by changing of centre of gravity. Moreover, the HL2 is equipped with a variety of sensors. These sensors include four visible light cameras, two infrared cameras, 1-MP Time-of-Flight depth sensor, an accelerometer, gyroscope, magnetometer and an 8-MP stills, 1080p30 video camera. The two infrared cameras are used for an eye-based rendering regarding tracking of eye-movement. [59] In



comparison with the HL1 the HL2 can track both hands instead of one, eye tracking has been added and the field of view is enhanced (34° versus 52°) [61]. Both the HL1 and HL2 are visualised in figure 1.



*Figure 1, left the HoloLens 1 and right the HoloLens 2*

### 3.4 Medical applications of the HoloLens

The HoloLens is used for different applications in different medical settings. Augmented reality is used in anaesthesia, neurosurgery, urology, oral and maxillofacial surgery, dermatology, plastic surgery and general surgery. [58], [62] In this section several examples of the HL used in clinical studies will be stated.

One of these research groups is the Meulstee group that is part of the 3D Lab of the Radboud University Medical Centre (Radboudumc). This group conducted multiple studies towards the use of AR in the operating room (OR). First, they conducted an accuracy study, to see if the HL was suitable for medical use. After this study they concluded that there is a large potential for AR in combination with accurate tracking systems for a wide variety of medical disciplines. [63] In a study on the use AR during craniomaxillofacial surgery, conducted by Thijs Bussink, they showed possibility for using AR during surgery. [55]

Kuzhagaliyev et al. [64] used the HL to determine the best location for needle placement in irreversible electroporation of pancreatic tumours. Kuhlemann et al. [65] researched the usage of the HL during endovascular interventions. Their goal was to replace X-ray with the HL to visualize the position and orientation of the catheter. Another medical field in which the use of HL was researched in the field of neurosurgery. Incekara et al. [66] used the HL to compare a virtual AR preoperative planning with the neuronavigation during brain tumour surgery. Van Doormaal et al. [67] used the HL as a stand-alone neuronavigation application in his study. These studies both show promising results regarding the use of AR in the OR.

The applications mentioned above all used different tracking and registration methods. Incekara used manual alignment as a registration method. [66] This means that the surgeon needs to correct the system manually every time the patient position is manipulated, leading to results that are highly user-dependant and are prone to human error. Kuhlemann on the contrary used landmark-based surface registration and a magnetic tracking system [65]. The biggest disadvantage of this technique is the need for an expensive external tracking system in the OR. Kuzhagaliyev used an optical navigation system that equipped with infrared cameras as an external tracking system [64]. Van Doormaal et al. [67] applied point-based registration, using a pointer with a quick response (QR) coded marker attached to it to mark special placed fiducials. These fiducials are placed on the fixated head of the patient. The fixation obviates the need for a tracking method. A disadvantage of this technique is the need of fiducials or other markers in sight during the surgery. To conclude, every registration and tracking method has its own advantages and disadvantages.

## 4 Development of the AR surgical navigation application

In this chapter, the workflow from the preoperative MRI to AR visualisation with the HoloLens 2 is presented. This chapter starts with a section that provides an overview of the process and which explains the various coordinate systems that are at stake. In the following sections, the three stages of which the process consists is explained.

### 4.1 Overview

This surgical navigation application depends on the following devices: the MRI-scanner, the HoloLens 2, a surgical pointer equipped with a QR-coded marker, and a QR-coded marker that can be attached to the patient. A QR-marker is a visual marker, the location of which can be measured by the HoloLens application.

Figure 2 provides an overview of the workflow. This workflow starts with the preoperative stage (green part). First, an MRI scan is made, as stated in section 2.3. Subsequently, the segmentation protocol, as stated in section 3.1, is used to create 3D meshes of the kidney, tumour and critical vessels. Manually, anatomical landmarks are selected and located. After that, the meshes are exported to the HoloLens.

The purpose of the second stage (blue part) is to find the geometrical transformations that are needed to have the preoperative MRI-based meshes correspond with real patient's anatomy, so that the HoloLens will be able to project the virtual objects onto the physical objects.

The first two stages are preparatory steps for the third stage. In this last stage (red part), movements of the HoloLens and of the patient are continuously tracked so that the HoloLens will be able to continuously maintain the correct overlay.

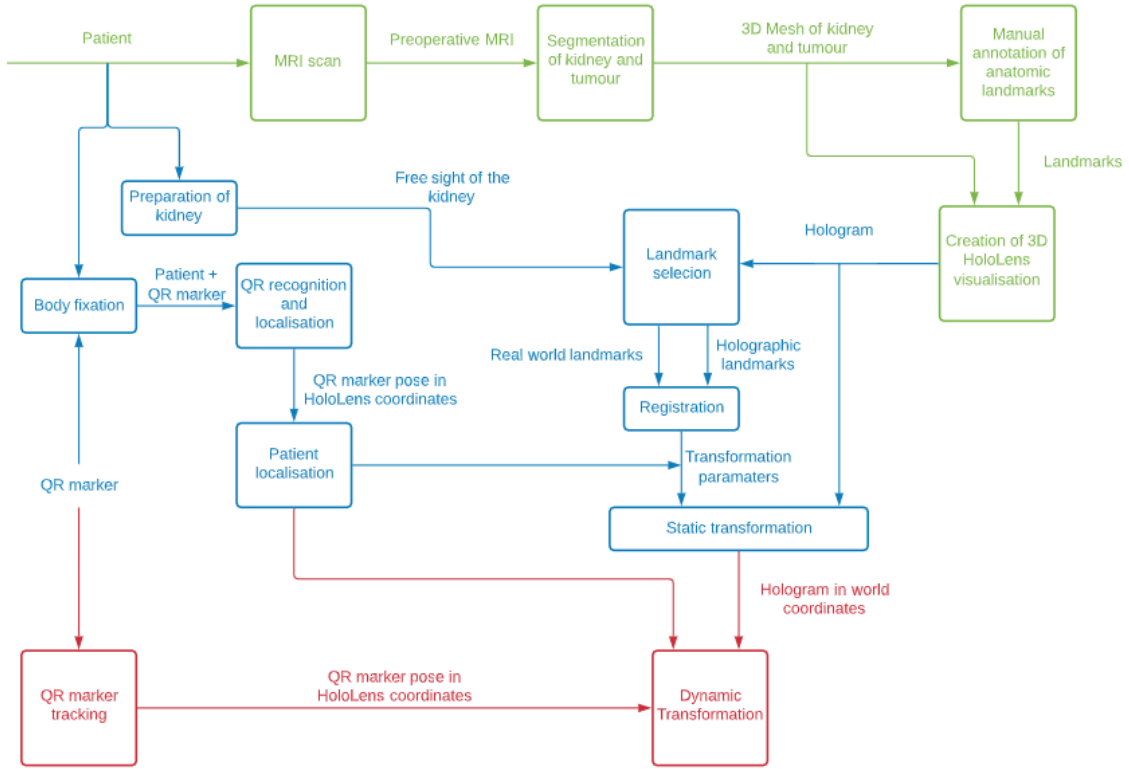


Figure 2, overview of the proposed workflow.

In this process, there are several coordinate systems at stake:

*MRI* : coordinate system that is defined by the MRI scanner

*H* : coordinate system, internally used by the HoloLens

*QP* : coordinate system that is associated with the marker that is attached to the pointer

*QR* : coordinate system that is associated with the marker that is attached to the patient

The relations between these coordinate systems are shown in figure 3. Transformations between the different systems are denoted by the symbol  ${}^wT_{from}$ . These  $4 \times 4$  matrices transform the representation of points from one domain to another. For instance,  ${}^HT_{MRI}$  transforms the representations of the points in *MRI* coordinates to a representation in *H* coordinates. Important to know is that if such transformation is known, the reverse transformation is also known since, for instance,  ${}^{MRI}T_H = {}^HT_{MRI}^{-1}$ .

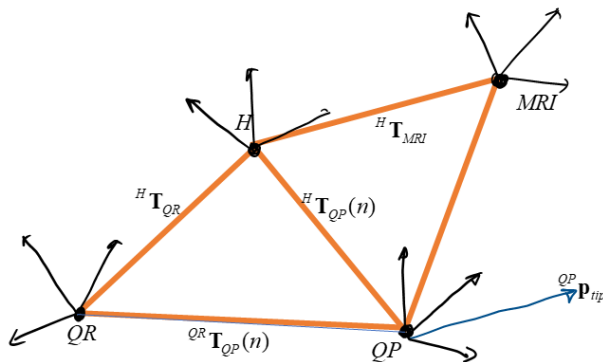


Figure 3, the various coordinate systems and the transformations between them.

## 4.2 Design options

In chapter 3.4 ‘Medical applications of the HoloLens’, we stated that there is a wide variation of possible registration and tracking methods for AR. In this section, we will discuss this variation and justify the chosen registration and tracking method for our AR surgical navigation system.

### 4.2.1 Registration methods

Registration is the alignment of one coordinate system with another. In the current application, this has to be done at several places: from the MRI to the HoloLens, and from the Physical world to the HoloLens. Several registration methods exist.

The first registration method that will be discussed is a manual overlay. Here, user has to rotate and shift the 3D visualisation manually until it fits with the real world. The advantages of a manual overlay are low level of entry, no need of complicated planning, and easy implementation due to the low technical needs. However, there are a couple of disadvantages to this technique. The first disadvantage is the reliance of the user, which makes it very prone to human errors. Moreover, it also has a low accuracy and is time consuming. [68], [69]

The second registration method is the use of a point-based registration, or landmark-based registration. During this registration method the user will pinpoint multiple, on beforehand defined, landmarks perioperative. The pinpointed landmarks will be matched with the on beforehand defined landmarks to calculate the rotation and translation matrices. The advantages of this technique are user friendliness, quickness to perform and its semi-automatic operational mode. However, this technique becomes invalid as soon as the camera or the tissue moves. [60], [68], [70]

Another possible registration method is the use of artificial markers. There are various markers which can be used, e.g. QR-markers, infrared markers or fluorescent markers. There are two methods to use markers in a registration method. The first one is to link the 3D segmentation to a marker. Thereafter, the marker has to be placed in the right position to have a sufficient overlay. The other method is to place internal markers inside the patient before the preoperative MRI. These markers are visible on the MRI, and can therefore be segmented. The segmented markers can be loaded into the HoloLens. The markers will remain in the patient until surgery. These markers are recognised intraoperative by the HoloLens. Due to the recognition and the fact that the markers are positioned at the same location as during the MRI, the overlay can be created. These markers are fast, accurate and robust. [11], [50], [68] However, it is not desirable for the patients to undergo an additional procedure.

Based on these advantages and disadvantages, and the fact that there is already experience with a combination with point-based registration and QR-markers, we choose to use this combination as the registration method.

### 4.2.2 Tracking methods

The first tracking method that will be discussed is the use of Vuforia (v9.1, PTC Inc. Massachusetts, USA). This is a fast and accurate, software development kit that can be used for both tracking and registration. Vuforia uses the visual output of the HoloLens’ build in cameras, to recognize 2D or 3D objects. Therefore, no additional hardware devices are needed to use Vuforia as the tracking device. However, the use of the HoloLens’ cameras has the downside of the need of a clear line of sight. [71]

Another possible tracking method is the use of Electromagnetic (EM) tracking. This tracking method is not depending on a clear line of sight, on the contrary of Vuforia. EM tracking is already used inside

the OR reaching an accuracy of around 0.5 millimeters. However, this technique has multiple disadvantages. It needs an additional external device; it requires magnetic responders in the target organ and one of the biggest disadvantages is distortion of this technique caused by metal tools. [69], [72] Especially the latter is a big disadvantage due to the use of the Thompson Bariatric System during NSS. [73]

Based on these advantages and disadvantages, and the fact that the combination of Vuforia and QR-markers registration have been used in other studies conducted by the 3D Lab of the Radboudumc [55], [63], these techniques will also be used for this study.

### 4.3 The preoperative stage

As stated in chapters 2.3 and 3.1, patients suspected with a WT get a preoperative MRI scan, which is segmented using 3D Slicer. These 3D segmentations consist of the kidney, main renal arteries and the tumour. These are loaded into the HoloLens using Unity software (v.2019.4.3f1, Unity Technologies, San Francisco) as 3D mesh objects with a MRI reference frame,  ${}^{MRI}\mathbf{M}_i$  with  $i = 1, 2, \dots$ .

Preoperatively, also a number  $N$  of identifiable virtual anatomical landmarks are selected in the MRI data. These virtual landmarks are denoted by  ${}^{MRI}\hat{\mathbf{I}}(n)$  with  $n = 1, \dots, N$ . The landmarks must be visible intraoperatively. However, preoperatively, the orientation of the kidney is unknown during NSS, and thus which physical landmarks are visible. Therefore, it is important to select a wide variation of virtual landmarks preoperatively, to give the surgeon the ability to choose the best set of landmarks intraoperatively.

The Unity software converts the MRI coordinates to HoloLens coordinates, but the transformation to do so can be freely chosen. In our implementation, arbitrarily, the identity transform  ${}^H\mathbf{T}_{MRI} = \mathbf{I}_{4 \times 4}$  was chosen so that the HoloLens representation of coordinates correspond with the MRI coordinates.

That is,  ${}^H\hat{\mathbf{I}}(n) = {}^{MRI}\hat{\mathbf{I}}(n)$  and  ${}^H\mathbf{M}_i = {}^{MRI}\mathbf{M}_i$

### 4.4 Patient registration

The second stage is patient registration. After registration, all data will be represented in the same coordinate system so that in the surgeon's view the MRI meshes overlay with the real anatomy. As stated before, the registration method of this application consists of a combination of a point-based method and localization of QR-coded markers. The renal orientation during the tumour resection is not preoperatively predictable. Therefore, intraoperative visibility of all the predefined MRI-based anatomical landmarks  ${}^H\hat{\mathbf{I}}(n)$  cannot be guaranteed. For this reason, the surgeon selects only the subset of landmarks that are visible from his point of view.

The visible virtual anatomical landmarks are defined into the real world using a surgical pointer. This surgical pointer is equipped with a QR-coded marker, and is therefore locatable by the HoloLens in HoloLens coordinates:  ${}^H\mathbf{T}_{QP}(n)$ . At the same time, the HoloLens is also able to measure the pose of the reference QR-marker:  ${}^H\mathbf{T}_{QR}$ . We assume that during registration, the patient does not move. Since the reference QR-marker is rigidly attached to the patient, it does not move either. Since the HoloLens can measure the pose of both QR-markers, physical anatomical landmarks of the patient can be expressed in the reference QR coordinates. [74]

The registration of the physical anatomical landmarks with the virtual anatomical landmarks consists of three steps. Firstly, determination of the physical anatomical landmarks in HoloLens coordinates, and subsequently in reference QR coordinates. Secondly, matching the physical anatomical landmarks with the predefined virtual anatomical landmarks. And lastly, determination of the transform between MRI coordinates and the HoloLens coordinates using the Procrustes algorithm.

#### Step 1: Determination of the locations of physical anatomical landmarks

The QR-code recognition is performed using Vuforia. When the holographic surgical application is started, Vuforia is started as well. The transform matrix of the QR-marker is updated for every frame and the HoloLens 2 gives an experience of 60 frames per second [59]. When the tooltip of the surgical pointer is located accurately onto a selected physical anatomical landmark, this landmark can be recorded. The poses of the reference QR-marker and the QR-marker, attached to the pointer, are given by the transformation matrices  ${}^H\mathbf{T}_{QR}(m)$  and  ${}^H\mathbf{T}_{QP}(m)$ , respectively. Here,  $m$  is the index of the  $m$ -th anatomical landmark.

The locations of the physical anatomical landmarks are determined by the location of the tip of the surgical pointer. The location of the tooltip is directly linked to the location of the QR-marker attached to the surgical pointer, and is represented by  ${}^{QP}\mathbf{p}_{tip}$ . Using the position of the tooltip and the transformation matrices, the location of the anatomical landmarks, expressed in the QR references coordinates, is:

$$\begin{aligned} {}^{QR}\mathbf{I}(m) &= {}^{QR}\mathbf{T}_{QP}(m) {}^{QP}\mathbf{p}_{tip} \\ &= {}^{QR}\mathbf{T}_H {}^H\mathbf{T}_{QP}(m) {}^{QP}\mathbf{p}_{tip} \end{aligned} \quad (1)$$

In this equation, the anatomical landmarks are enumerated  $m = 1, \dots, M$ .

#### Step 2: Landmark association

After the determination of all physical anatomical landmarks, it is important to match these landmarks with the associated virtual landmarks. There are multiple methods to execute this matching process. The first possibility is to instruct the surgeon to acquire the physical anatomical landmarks in the same order as the predefined virtual anatomical landmarks. However, this requires additional effort of the surgeon as he/she will have to remember the exact order of the virtual anatomical landmarks. Moreover, preoperatively it cannot be guaranteed which landmarks will be visible intraoperatively, and which order will be convenient for the surgeon.

To overcome these disadvantages, we implemented a brute-forcing method. First, note that  $M \leq N$ . Therefore, we consider every possible permutation of the set  ${}^H\hat{\mathbf{I}}(n)$  consisting of  $M$  landmarks, and associate that with the ordered set of  $M$  landmarks  ${}^{QR}\mathbf{I}(m)$ . To each permutation, the Procrustes algorithm is applied. The permutation with the lowest root-mean-square error (RMSE) will be used as the right one. Note that a set of  $N$  landmarks has  $N!/(N-M)!$  permutations of  $M$  landmarks. Note that this number of permutations is soon excessively large. For instance, with  $N = 10$  and  $M = 5$ , the number of permutations is 30240, but with  $N = 15$  and  $M = 5$ , this number becomes around  $10^{10}$ .

### Step 3: Alignment of coordinate systems; Procrustes algorithm

Each possible association in step 2 should be validated by applying Procrustes algorithm. Such an association consists of the ordered set  ${}^{QR}\mathbf{I}(m)$  and an ordered subset from  ${}^H\hat{\mathbf{I}}(n)$  which form the association  ${}^{QR}\mathbf{I}(m) \square {}^H\hat{\mathbf{I}}(n_m)$ .

The Procrustes algorithm consists of three transformations, translation, rotation and uniform scaling. However, scaling is not used in this application. The alignment is defined by a 3D rotation and a 3D translation. Mathematically, this is represented by a rotation matrix  $\mathbf{R}$  and a translation vector  $\mathbf{t}$ :

$${}^{QR}\hat{\mathbf{I}}(n_m) = \mathbf{R} {}^H\hat{\mathbf{I}}(n_m) + \mathbf{t} \quad (2)$$

Taken together, the rotation matrix  $\mathbf{R}$  and translation vector  $\mathbf{t}$  are embedded in the transformation matrix  ${}^{QR}\hat{\mathbf{T}}_{MRI}$ . The Procrustes algorithm determines  $\mathbf{R}$  and  $\mathbf{t}$  such that the root mean square error

$$RMS = \sqrt{\frac{1}{M} \sum_{m=1}^M \| {}^{QR}\mathbf{I}(m) - {}^{QR}\hat{\mathbf{I}}(n_m) \|^2}$$

is minimal. The details of the algorithm can be found in Paláncz et al. (2010). [75]

As described in step 2, the Procrustes algorithm is applied to each association, and the one with the minimal mean square error is deemed to be the correct one. This finally provides the transformation matrices  ${}^{QR}\mathbf{T}_{MRI}$ . Since  ${}^H\mathbf{T}_{MRI} = \mathbf{I}$ , this also provides  ${}^{QR}\mathbf{T}_H$  and  ${}^H\mathbf{T}_{QR}$ . These mappings allow the HoloLens to overlay the view of the 3D meshes  ${}^{MRI}\mathbf{M}_i$  from the MRI data with the physical world.

## 4.5 Tracking

In the third stage of the workflow, the movements of the HoloLens and the patient are compensated to maintain the correct overlay. For this purpose, the HoloLens measures the transformation matrix of the reference QR marker. Suppose that the transformation matrix as determined during registration in stage 2 is denoted by  ${}^{QR}\mathbf{T}_{MRI}$ . The transformation matrix at an arbitrary point in time  $t$  is denoted by  ${}^{QR}\mathbf{T}_H(t)$ . Then, the mapping of the MRI meshes becomes

$${}^{QR}\mathbf{T}_H(t) {}^H\mathbf{T}_{QR}(0) {}^{QR}\mathbf{T}_{MRI} {}^{MRI}\mathbf{M}_i \quad (1)$$

## 4.6 Realization

This application has been developed with Unity (v2019.4.3.f1, Unity Technologies, San Francisco). This is a cross-platform game engine, which can be used for developing 2D, 3D, VR and AR applications for a wide variety of devices. For this thesis a 3D AR application for the HoloLens 2 was built. This application consists of multiple scripts, which are written in C#. Microsoft Visual Studio (2019 version 16.4.0, Washington, USA) was used as an integrated development environment for the development of these scripts. Two software-development kits, Vuforia Augmented Reality SDK, and the ONS SDK, were integrated into Unity. The first SDK is used for recognition and tracking of 2D and 3D objects in real time using the integrated cameras of the HoloLens 2. The latter SDK consists of multiple codes and building blocks and is created by the 3D Lab of the Radboudumc. This is used as the starting point of this application.

### 4.7 Clinical workflow

The clinical workflow starts after creation of patient-specific 3D models through the technique described in section 3.1. The resulting kidney and tumour 3D meshes are loaded into Thereafter, the virtual anatomical landmarks are determined, by combining the virtual landmarks and the 3D mesh the 3D kidney visualisation is created for the HoloLens. The visualisation is loaded into the holographic surgical navigation application.

Before this application is started, the user is obligated to run the build-in eye calibration of the HoloLens. During the eye calibration the HoloLens measures the eye position of the user to render the holographic world. By using the eye position the best hologram positioning is ensured for the specific user. [59]

When the eye calibration is completed, the holographic surgical navigation application can be started. This application consists of three main scenes, namely the main menu, the choosing landmarks menu and the registration and tracking scene. The main menu consists of two different button sets. The first set of buttons enables the possibility to choose the right patient. The other button set consists of one button to close the application. The main menu is visualised in figure 4.

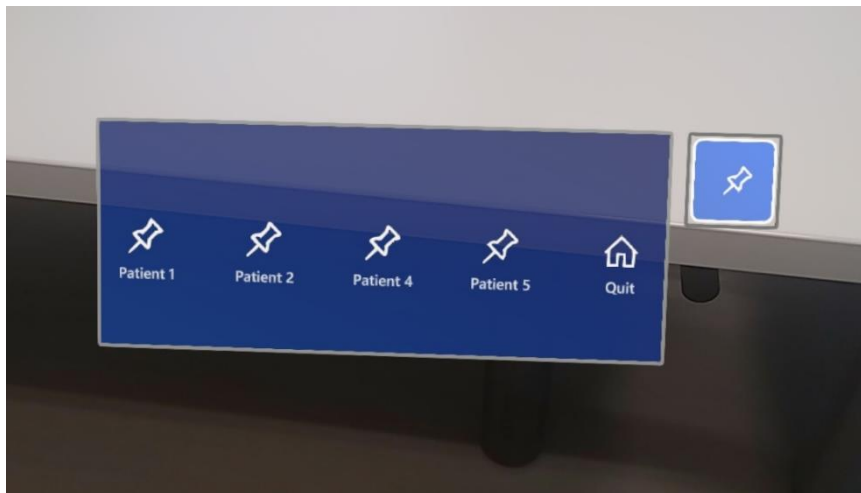


Figure 4, the main menu of the home-build HoloLens 2 application.

By choosing one of the “patient” buttons, the user is guided to the menu where he or she can choose the virtual anatomical landmarks of the specific patient. By opening this menu, the 3D kidney model is visualised. Besides the kidney itself, this model consists of the tumour, aorta, renal artery, renal vascularization and the predefined virtual anatomical landmarks. The user can enlarge, rotate and translate this 3D model in the HoloLens coordinate system. Besides this 3D model, eight buttons are visualised, of which five have the same function. These five buttons are used to enable and disable the predefined virtual anatomical landmarks. The chosen landmarks are stored in a new dataset which will be used during registration and tracking. However, it is not possible to delete a specific landmark out of this new dataset. Meaning, if the user stored an incorrect landmark, he has to use the retry button, this button will clear the new dataset. The last two buttons have an opposite goal. Namely, one button will guide the user back to the main menu and the other will guide the user further to the registration and tracking scene. The patient scene is visualised in figure 5.



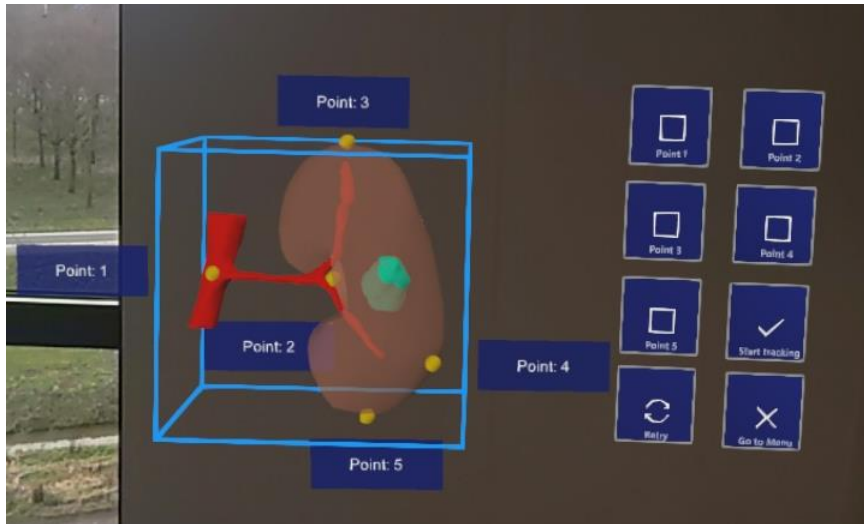


Figure 5: visualisation of the patient scene.

The third scene is the registration and tracking scene. In this scene the chosen virtual landmarks in the latter scene have to be pinpointed in the real world. These points can be pinpointed using the surgical pointer. When the surgical pointer is positioned onto one of the physical landmarks, the anatomical landmark can be fixated with two methods. First, the voice command “fix points” can be given by the user. The other method is to tap onto the QR-marker attached onto the surgical pointer. If the dataset of the physical anatomical landmarks reached the same size as the chosen virtual landmarks the overlay is projected using the Procrustes algorithm. If the overlay is not perfectly projected onto the kidney, or if an anatomical landmark was placed incorrectly, the user can retry the registration procedure by pressing the retry button. Other buttons integrated into this scene are the “go to set landmarks” (previous scene), “go to the main menu” and the “quit” button to close the application. All four buttons are linked to the reference QR-marker, to ensure their locations in the HoloLens coordinate system. The entire clinical workflow is visualised in figure 2.

## 5 Part 1: Accuracy of QR-coded marker recognition

### 5.1 Introduction

For an optimal tracking system it is preferable to fixate the reference QR-coded marker onto the target organ. If this is applied, the system can correct for movements of the target organ. However, during the studies of the 3D Lab of the Radboudumc, all used QR-coded markers had a size of 5x5 cm. This dimension is too large for fixation onto a children's kidney. Because it will cover-up most of the kidney and the tumour, which will cause a hindered view of the surgical site. Therefore, it is desirable to decrease the size of the QR-markers.

Researchers of the Radboudumc tested the ability of the HoloLens to recognise QR-markers in different locations of the field of view. [55], [63], [76] However, this is executed with the HL1. It is expected that the HoloLens 2 has an increased ability to detect QR-codes, due to the technical improvements. Consequently, it could be possible to decrease the marker size. This is preferable, to decrease the degree of hinderance for the surgeon.

This part of the study is designed to determine the minimal size of the QR-marker, without reduction of the accuracy and trackability of the markers. It is important that the QR-marker is tracked during surgery, so the hologram is visual all the time. Besides the capability to track the marker, it is also important to have an accurate determination of the location. The visualisation of the tumour is eventually the most important feature of this algorithm.

It is important to know how accurate the registration of the QR-marker is under different viewing angles. Therefore, the angle detection will be studied additional to the location detection of the QR-marker. This is important for the detection of the QR-marker which is fixated onto the surgical pointer. During the pinpointing of the physical anatomical landmarks, the surgeon rotates the surgical pointer to reach the optimal landmark location. This measurement results in a guideline for the orientation of the surgical tool and of the reference QR-marker.

In summary, the goal of this part of the study, is to investigate the minimal size of the QR-marker, without reduction of the accuracy and trackability. This will be investigated by the detection of differences in accuracy of the detection at different positions and different the angles.

### 5.2 Materials and methods

For this part of the study a measurement setup is built of Lego. Lego was chosen because of the reproducibility. Lego has specific dimensions and by clicking the blocks the exact same location can be achieved over time. To assure a stable HoloLens coordinate system, the HoloLens is placed onto a 3D printed bearer, which was fixated onto the table. The QR-marker was fixated onto a Lego tower with a specified height to ensure that the QR-marker was perpendicular to the line of sight of the HoloLens cameras. The tower was placed onto a Lego ground plate at a distance of 50 cm of the HoloLens, because this corresponds with the distances between the HoloLens and the kidney during surgery. To ensure the reproducibility of this study, a Lego track was made.

This track ensures the same movement of the QR-marker during all measurements. It consists of different locations, where the tower has to be fixated onto the ground plate. These locations are indicated by Lego blocks which are placed perpendicular to the track. These stop blocks indicated four different locations in a horizontal line and five different locations to simulate a movement across the line of sight.

There are several steps performed by the observer to execute these measurements. The first step was to switch on the HoloLens and to connect it with the laptop using the Windows device portal. The connection with the device portal was essential for starting the application on the HoloLens. Moreover, the device portal had the ability to give a live visualisation of the HoloLens recording. This was useful to see if the QR-marker was still being tracked. After starting the application and the live recording visualisation, the observer also started a stopwatch. During the measurements the observer was responsible for the manual movement of the QR-marker, every 30 seconds. It was important move the QR-marker as smoothly as possible. The HoloLens can experience loss of tracking when the QR-marker is replaced too quickly. The QR-marker will firstly be moved to the different locations on the right, than it will be placed back to the starting position. Thereafter, it was moved closer to the HoloLens and finally it will be moved further away.

For the calculation of the angles between the QR-marker and the HoloLens a comparable setup is used. The HoloLens is still placed onto the same bearer, and the same Lego tower is used to fixate the QR-markers on. To ensure that the QR-markers rotate comparable during the measurements, a rotation compass. This compass is drawn onto the surface of the table. It was created from the middle point of the QR-marker, to ensure that it is easy to use, the outline of the tower is drawn for every rotation. During this measurement the QR-marker will rotate every 30 seconds with 10 degrees till 80 degrees.

The above explained measurements are performed ten times per QR-marker dimension. During this part of the study, we used three different dimensions of QR-markers, 5x5 cm, 4.5x4.5 cm, and 4x4 cm.

For these measurements an additive application is created. This application uses the same tracking method as the final application. However, it has an additional script which allows the system to create a text file onto the HoloLens memory. This text file will be filled with HL coordinates of the QR-markers and their rotation in quaternions. If the QR-marker is not detected the text “not found” will be placed into the text file instead of the coordinates. These coordinates will be updated every five seconds. Another difference with the final application is the used Vuforia tracking mode. During this measurement the extend-tracking is turned off. This tracking state of Vuforia provides the target’s pose information, even if the target is not directly detected. [77] For this measurement it is important that the tracking is performed based on the direct tracking of the QR-marker, and not based on assumptions of the location of the QR-marker.

The received data will be evaluated with windows Excel (365, Microsoft, Washington, USA) and SPSS (IBM Corp. IBM SPSS Statistics for Windows, version 25.0. Armonk, NY: IBM Corp.). First all data will be gathered in Excel. The mean X or Z coordinates, depending on the movement, of two consecutive locations will be subtracted. These differences represent the distance between the two locations. These distances will be compared with the real distance, which is 4.74 cm. Thereafter, SPSS will be used to evaluate the differences of the measured distances between the different QR-markers dimensions, and between the different locations. A paired-sample T test is used to determine these differences. The paired-sample T test will be executed for both evaluating the measurements within the same QR-marker dimensions as between the different dimensions. The evaluation between the different dimensions will be performed on the same movements. The differences between the detection of rotation of the QR-markers will also be evaluated with the paired-sample T test between the QR-marker dimensions.

### 5.3 Results

In this section the results of the measurements explained in section 5.2 will be visualised. This section is composed as follows; first the results of the 5x5 cm QR-marker will be visualised, thereafter the 4.5x4.5 cm QR-marker and the 4x4 QR-marker, finally the comparison between the different QR-marker dimensions will be visualised. After the results of the detection of movement of the QR-markers is visualised, the results of the rotation is shown.

The mean values of the 5x5 cm QR-marker categorized on the locations is given in table 1. The results of the comparison between the different X-positions and different Z-positions results in three statistical significant outcomes. These statistical differences are measured for the differences between the first and fourth, second and third, and the second and fourth movements in Z direction. The P values are 0.009, 0.033 and 0.001 respectively.

*Table 1: The mean distances between two consecutive positions of the QR-marker, with dimensions of 5x5 cm, in centimetres and compared with the real distance of 4.74 centimetre.*

Position	Mean value [cm] (std.)	Error [cm]
Distance between X-position one and two	4.69 (2.1)	0.05
Distance between X-position two and three	4.85 (1.9)	0.11
Distance between X-position three and four	4.65 (2.0)	0.09
Distance between Z-position one and two	4.70 (2.8)	0.04
Distance between Z-position two and three	4.53 (2.2)	0.21
Distance between Z-position one and four	4.96 (4.9)	0.22
Distance between Z-position four and five	4.96 (1.8)	0.22

The results of the 4.5x4.5 cm QR-marker categorized on the locations are shown in table 2. The results of the comparison between the different X-positions and different Z-positions gives two statistical significant outcomes. These statistical differences are measured for the differences between the second and third, and the second and fourth movement in Z direction. The P values are 0.017 and 0.020 respectively.

Table 2: The mean distances between two consecutive positions of the QR-marker, with dimensions of 4.5x4.5 cm, in centimetres and compared with the real distance of 4.74 centimetre.

Position	Mean value [cm] (std.)	Mean error [cm]
Distance between X-position one and two	5.25 (2.5)	0.51
Distance between X-position two and three	5.27 (3.2)	0.53
Distance between X-position three and four	4.91 (6.5)	0.17
Distance between Z-position one and two	5.19 (7.1)	0.45
Distance between Z-position two and three	4.85 (4.3)	0.11
Distance between Z-position one and four	5.40 (6.5)	0.66
Distance between Z-position four and five	5.19 (3.0)	0.45

The mean values of the 4x4 cm QR-marker categorized on the locations are shown in table 3. No statistical significant outcomes are measured between the different X-positions and Z-positions.

Table 3: The mean distances between two consecutive positions of the QR-marker, with dimensions of 4x4 cm, in centimetres and compared with the real distance of 4.74 centimetre.

Position	Mean value [cm] (std.)	Mean error [cm]
Distance between X-position one and two	5.88 (2.9)	1.14
Distance between X-position two and three	5.98 (2.1)	1.24
Distance between X-position three and four	5.82 (5.7)	1.08
Distance between Z-position one and two	5.84 (3.3)	1.10
Distance between Z-position two and three	5.98 (2.9)	1.24
Distance between Z-position one and four	6.71 (14.4)	1.97
Distance between Z-position four and five	5.86 (2.8)	1.12

Comparing the outcomes of the different QR-marker dimensions, practically all different positions are statistically significant, the third x movement between the 5x5 and 4.5x4.5 marker, first z movement between 5x5 and 4.5x4.5 marker, and the third z movement between the 5x5 and 4.5x4.5 and 5x5 compared to the 4x4 markers excepted. The p-values of these not statistical significant differences are 0.212, 0.91, 0.165 and 0,068 respectively. The mean distances of the different QR-marker dimensions are visualised in figure 6.

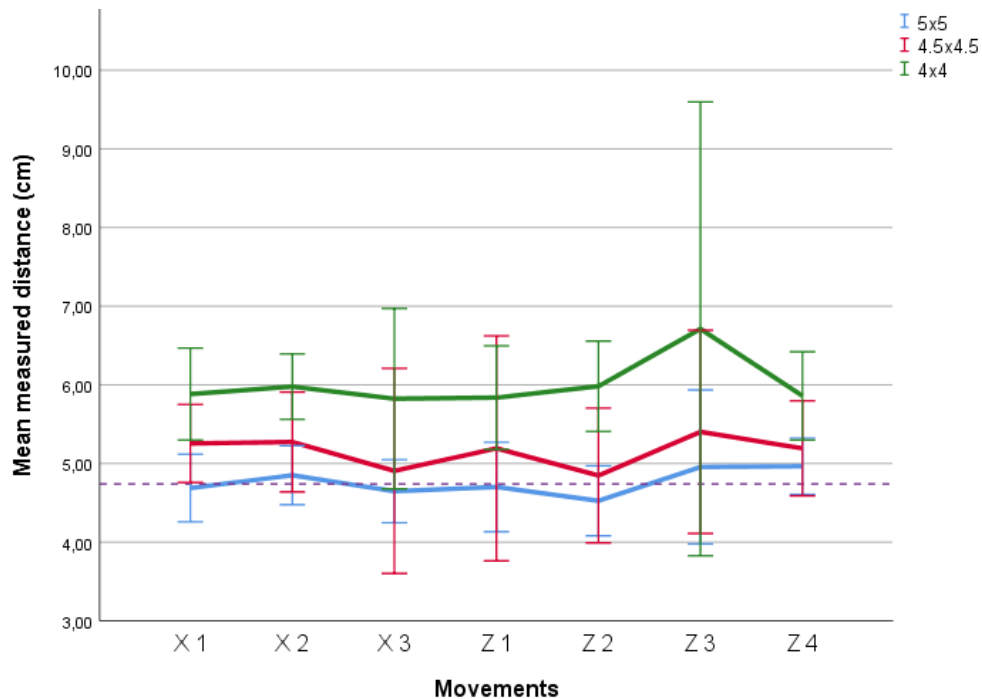


Figure 6, mean measured distance of the different QR-marker dimensions per position and the standard deviation per movement. The purple dashed line represents the real distance.

The mean measured angles of the different rotations per QR-marker dimension is visualised in table 4. The HoloLens was not able to detect the QR-marker with an angle of 80 degrees. The paired-sampled T test shows that almost all differences between the QR-marker dimensions are statistically significant, except the differences between the 4.5x4.5 cm and 4x4 cm marker at a rotation of zero degrees, the 5x5 cm and 4x4 cm marker at 10 degrees, 20 degrees, and 30 degrees. These comparisons had a p-value of 0.155, 0.170, 0.371 and 0.670 respectively.

Table 4: the mean values of the measured angle of the different QR-markers, measured in degrees.

Degrees	0	10	20	30	40	50	60	70	80	Mean error
5x5	2.24	11.71	21.17	31.34	40.58	50.76	60.37	69.80	-	1.04
4.5x4.5	3.68	14.14	23.85	33.47	43.88	53.70	62.92	71.77	-	3.43
4x4	4.33	10.98	21.15	31.25	41.62	51.51	61.18	70.29	-	1.54

## 5.4 Discussion

The differences of the 5x5 cm QR-marker were significant between measurements of different position in the Z-direction. These differences combined with the mean error of the positions, suggests that the HoloLens 2 identified the QR-marker optimally at a distance of 50 cm. The mean error of this distance is less than a millimetre. The other two dimensions showed a worse outcome at a greater distance as well.

The results show significant differences between the three QR-marker dimensions. The 5x5 cm QR-marker showed the best accuracy. This means it is not desirable to use a smaller QR-marker at a distance of 50 cm of the HoloLens. Consequently, it is not possible to attach the QR-marker onto the kidney. 5x5 cm will hinder the surgical site, making the surgical removal of the tumour impossible. For this matter the location of the reference QR-marker has to be determined again.

Accuracy consists of two parts, precision and trueness. Precision is agreement of the different measurements, trueness is the agreement of the measurement and the real value, according to the International Organisation of Standardization. [78] The precision can be determined with the standard deviation, and the trueness with the mean error. From the results it can be concluded that the 5x5 cm QR-marker has both a better trueness as precision than the other two dimensions.

The results in the detection of the rotational angle show a mean error of 1.04, 3.43 and 1.54 degrees for the three dimensions. There is no logical explanation for the increased accuracy of the 4x4 cm dimension compared to the 4.5x4.5 cm QR-marker. Possible explanations are, the QR-marker was not attached optimal on the Lego tower, or the QR-marker was not printed optimal. However, with the second explanation it is expected to see the same trend in the detection of different positions of the QR-markers, while the same QR-markers are used. Another possibility is changed lighting conditions. The measurements are performed on different days. The weather conditions could influence the measurement results.

These measurements are executed in a room with two small windows and one artificial light source. The weather conditions were quite similar during the measurements. However, the HoloLens needs the right lighting conditions to work optimally. [79] The lighting conditions in the room were not optimal. Suggesting, the errors could be smaller, when measuring in optimal conditions. The lighting conditions of the operation room are of a superior quality compared to this measurement room. Therefore, it is presumably to have an increased ability of QR-coded marker recognition in the operation room. The conditions were comparable between the different measurements, so the differences between the QR-marker dimensions would probably not differ.

All three dimensions show a decrease in accuracy when the QR-marker is moved away from the HoloLens. This suggests that the optimal registration and tracking distance is around 50 cm of the HoloLens camera. This corresponds with the findings of Putta (2019) [80]. This leads to the recommendation for the surgeons to hold the surgical pointer in a not fully extended arm. [81] In terms of rotation the only limit the surgeons will experience is the maximal rotation of 70 degrees. The HoloLens was not able to recognise the QR-marker at a rotation larger than 70 degrees. This is desirable, because it is difficult to keep the surgical pointer at the same rotation during the surgery. This will improve the user-friendliness of the application.

### 5.5 Conclusion

The 5x5 cm QR-marker is the only dimension which is sufficient as a trackable object for this application. The detection of the QR-marker with different angles showed that the detection is properly till an angle of 70 degrees. Which makes it suitable to use as a tracking device for the surgical pointer during NSS. However, the 5x5 cm dimensions are too big to use as a tracking device onto the kidney itself. For this matter the location of the reference QR-marker has to be determined again. Concluding, this dimension will be used in the subsequently chapters of this thesis, for both the surgical pointer and the reference QR-marker.



## 6 Part 2: Accuracy based on a 3D-printed kidney phantom

### 6.1 Introduction

In the latter section we determined that the recognition of the QR-marker is sufficient as a registration and tracking method using the HoloLens 2. However, the accuracy is not dependent purely on the detection of the QR-marker. The application also depends on the user, meaning it is expected that the accuracy decreased compared to the accuracy of chapter 5. Therefore, after the determination of the accuracy of the QR-marker registration, the accuracy of the entire application needs to be determined. Moreover, the application should have a maximal error of 5 mm, not associated with dependency of different users.

There are multiple variables associated with the user. The first variable is the possibility of the HoloLens to run the eye-calibrate of the user. If this calibration is not performed optimally, the overlay of the surgical pointer will not be perfect. This results in the possibility that the physical anatomical landmarks are fixated onto the wrong positions, which results in an incorrect overlay of the 3D kidney hologram. Another variable is the ability to correctly interpretate the virtual landmarks. The user is responsible for the interpretation of these points and the fixation of the anatomical landmarks onto the patient.

This application should be suitable for all paediatric surgeons. Therefore, it is important to study the effect of these variables on the accuracy of this application. The accuracy of the application should be greater than 5 mm for different observers. This would indicate that the application is suitable for different users.

The goal of this chapter is to validate the self-developed application based on a phantom study conducted with 10 observers. With a positive outcome of this validation, a clinical validation may be performed. However, if the errors of the observers are greater than 5 mm, the application should be improved.

### 6.2 Materials and methods

For this section of the study a 3D printed kidney phantom made of white polylactic acid (PLA) is used. This phantom is created from the MRI scan of a real patient of the Princes Máxima Centre. This phantom is fixated onto a 3D printed bearer specially designed for this specific phantom, to ensure that the phantom will not move during the anatomical landmark registration. The surgical pointer is a 3D printed replica of the surgical 3D pointer used by the 3D Lab of the Radboudumc.

Before the real accuracy measurement begins, the observers have to finish multiple steps. The first step is to conduct the build in HoloLens eye-calibration. The second step is to get familiar with the HoloLens 2, because none of our observers had prior experience with the HoloLens. Therefore, the observers were asked to perform “tips” a build-in application for learning the basic gestures to control the HoloLens 2. After completing this application, the observer is asked to close all outstanding applications. Subsequently “Kidney registration and tracking” is started and the accuracy measurement begins.

This measurement is twofold. First the observer has to follow the steps as described in the workflow in chapter 4.4 until the final step. During this second part the observer is asked to mark highlighted points onto the phantom with a marker. There are six points that needs to be marked. These points are shown one at a time. The observer needs to select the next point if he marked the point correctly onto the phantom. This second step is shown in figure 7.

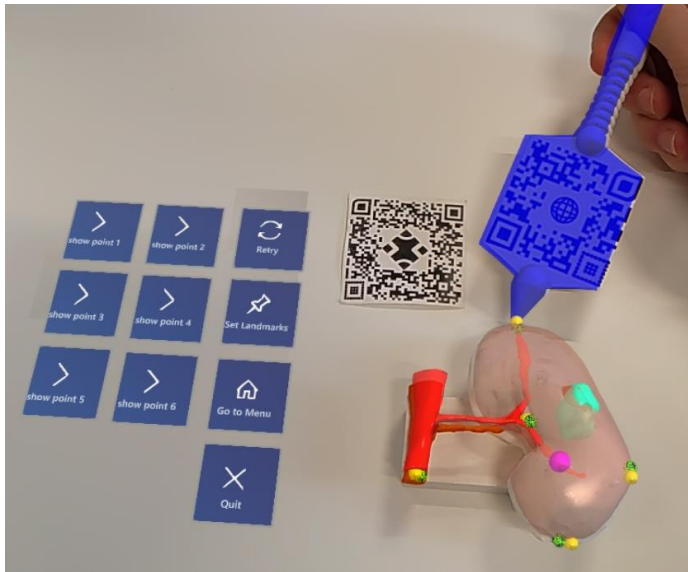


Figure 7, The holographic overlay of the kidney (salmon), tumour (cyan), artery (red), and a positional marker (purple) on top of the 3D printed model are shown. Additionally, the QR-marker used for tracking and the QR-marker placed onto the surgical pointer (blue) are visible. The surgical pointer is used to pinpoint the anatomical landmark for the HoloLens 2. There is a slight offset of the overlay due to a difference in height of the camera of the HoloLens 2 and the eyes of the user.

The marked points can be evaluated with a 3D printed calibration mold. This mold is specifically designed for this phantom. It fits specifically over the phantom in only one way. The mold consists of six holes placed corresponding with the actual position of the visualised points. These holes have the same diameter as a pencil, which makes it possible to set a mark on the right place. The distance between the two markings will be measured using a digital caliper (Sencys schuifmaat digitaal 15cm) with an accuracy of 0.01 mm. The distance between those markings will be measured by two different observers, the mean distance of both observers will be used as the error of the observer.

For this measurement 10 observers are asked to go through the described protocol. Three out of ten are paediatric surgeons, three are medical students, two observers are technical physician, and the last two observers are technical physicians graduation students.

All measurements were executed in the same location. This room is illuminated only with artificial light sources. This guaranteed the same circumstances during all measurements. To assure that other variables are the same for all observers the HoloLens will be fully charged before every measurement.

The data gained from this section will be evaluated with SPSS. The Shapiro-Wilk test is used to test the normality. The differences between the volunteers is determined with an one-way ANOVO combined with either the Turkey HSD when the data meets the assumption of homogeneity of variance or the Games Howell post hoc test when the data does not meet the assumption of homogeneity of variance. [82]

### 6.3 Results

The distance between the marked points by the observers and the real locations are visualised in figure 8. Unfortunately, one positional marker was excluded from the measurements due to a complete misalignment of the holographic marked position and actual position given by the calibration mold. Of these marked positions, one had an error of 0 mm. However, Three of the fifty marked positions were placed with a distance greater than the maximal clinical error of 5 mm of the actual position by two different observers. These three errors had a value of, 6.71, 7.41 and 11.65 mm. The corresponding boxplots of the observers is visualised in figure 9. The mean and standard deviation of all fifty errors combined are 3.16 and 1.82 mm respectively. Only one observer obtained a mean error above the clinically maximally allowed error of five mm, see table 5 .

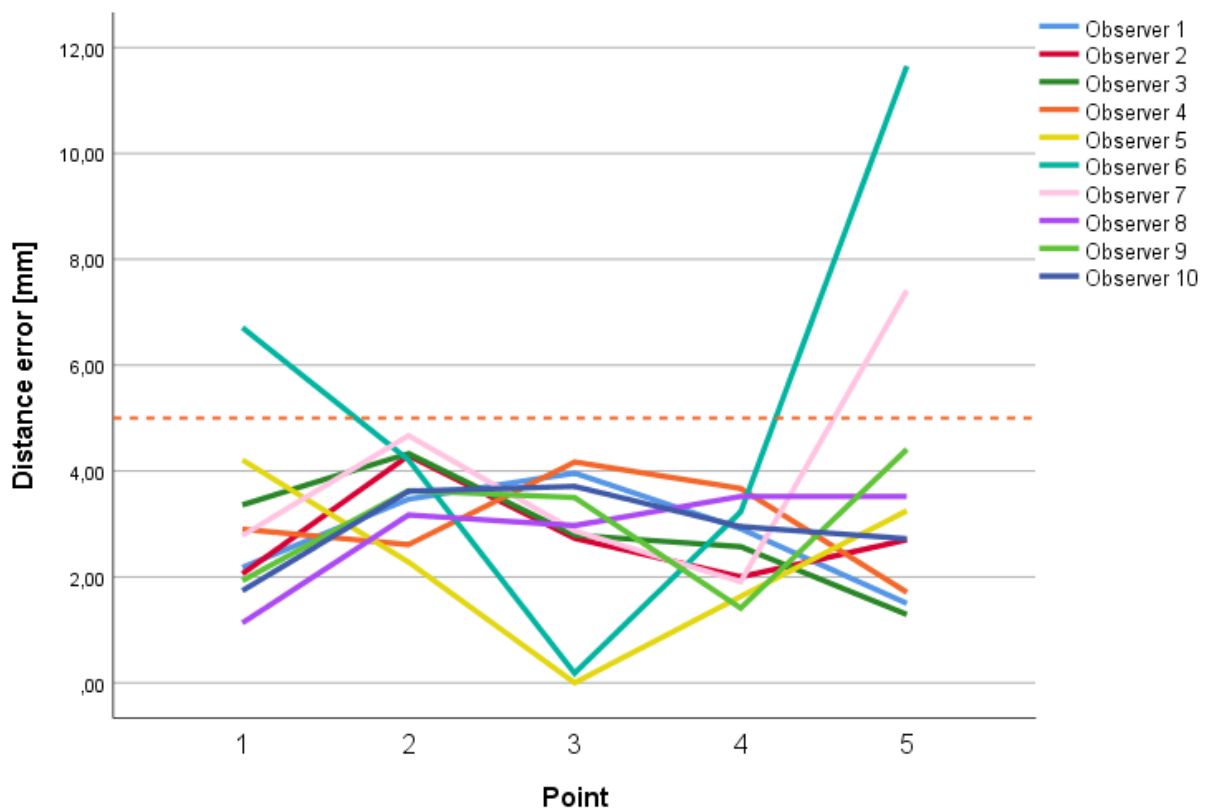


Figure 8, the distance error in millimetre of all observers per marked point. The orange dashed-line represents the clinically maximally allowed error of 5 millimetre.

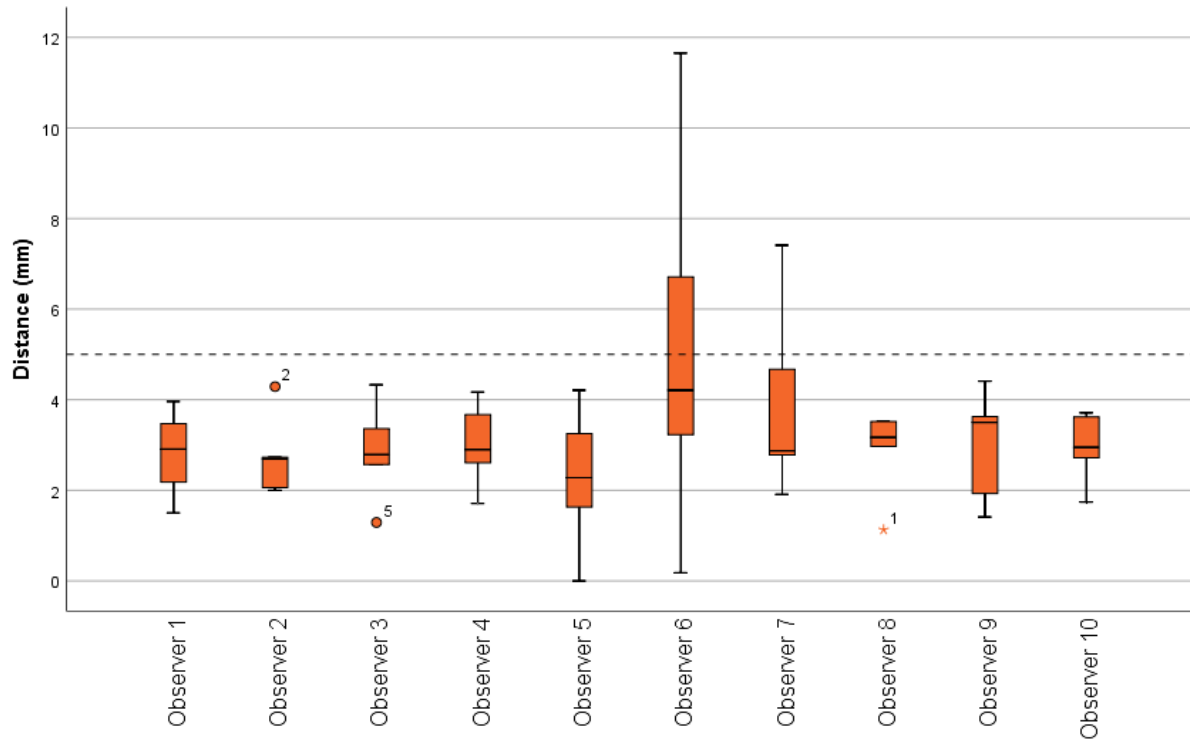


Figure 9, boxplots of all errors of all points combined per observer. This figure shows the clinically maximally allowed error (dashed line), and the variation per observer. Three points (1, 2, and 5) are significantly different of the other points set by the observer, however, these points all are within the clinically maximally allowed error.

Table 5, the mean error of the point per observer, with their standard deviation.

Observer	Mean error [mm] (Std.)
1	2.80 (0.99)
2	2.76 (0.92)
3	2.87 (1.11)
4	3.01 (0.95)
5	2.27 (1.60)
6	5.20 (4.30)
7	3.93 (2.19)
8	2.86 (1.00)
9	2.98 (1.26)
10	2.95 (0.80)
Mean of all observers	3.16 (1.82)

The differences between the observers are analysed in more depth, the Shapiro-Wilk test showed that not all datasets meets the assumption of homogeneity of variance, therefore the ANOVO test with a Games Howell post hoc test was performed. These tests showed no significant differences between the observers, with a p-value between 0.876 and 1.

Looking at the differences between the measured points by all observers no significant differences were found using the ANOVO test with a Games Howell post hoc test, with  $0.056 \leq p \leq 1$ .

The direction of the error of the observers did not have a systematic deviation. Most marked points at the cranial side of the kidney phantom had a deviation in the cranial lateral direction. While in general the points at the caudal side of the kidney showed a deviation with a medial caudal direction.

### 6.4 Discussion

The results of this section showed that the mean error of this system, used by unexperienced observers, has a mean accuracy of 3.16 mm with a standard deviation of 1.82 mm which is within the clinically determined limit of 5 mm. However, three of the fifty points exceeds the clinically maximally allowed error of 5 millimetres. Two of these three were set by the same observer. This observer notified that she could not reach the holographic highlighted point easily, due to a small reach. This is a possible explanation for this observer, as the two outliers were corresponding to the two most cranial points. The outlier of the other observer corresponds with the most cranial lateral location. It is suggested that the observer had a difficulty to interpret the right location of the highlighted point.

Looking further into the accuracy of this application, the mean error and the standard deviation are both sufficient for our specific clinical problem. Looking at the variation of deviation directions, we observed that the three cranial points have a cranial deviation, and the caudal points a caudal deviation. This suggests that the deviation is caused by the interpretation of the observers. The curvature of the kidney could play a role in this interpretation. The holographic highlighted points are visualised with a sphere. The observers were asked to mark the centre of these spheres with the marker. However, the position of this mark differs when the markers are held vertical or on a 90° angle with the curvature of the kidney. Most observers held the marker in a vertical position, which means that the marked points will shift in more caudal direction at the caudal part of the kidney due to the descending slope of the phantom. The same applies to the cranial points.

This study has multiple limitations. First, as stated in the previous paragraph the holographic highlighted points are visualised with a sphere. The diameter of this sphere is 7.5 millimetre. This diameter is chosen, for a good visibility for the observers. However, these spheres have a larger diameter compared to the point of the marker. Therefore, the observer needs to interpretate the centre of the spheres themselves. This creates an extra variable for the accuracy of the application. This extra variable results in an increase in the error measured by the observers, and thus a decrease in the accuracy.

Secondly, the observers were asked to set a single point with the marker. However, due to the holographic overlay the observers could not see if the marked point was visible on the phantom. To ensure a good marking, most observers gave more pressure on the marker as necessary, resulting in a larger marking. This creates more deviation between the two measurers. The two measurers measured the distance between the centre of the marked point and the centre of the real position, marked with the mold. The differences between the measurers increased, when the size of the marked point increases.

Another limitation of this measurement is the use of a rigid phantom. This is not representable for a real kidney. Therefore, we cannot directly correlate this measured accuracy with the accuracy obtained in a clinical setting during nephron-sparing surgery. The real kidney is deformable, and thus more challenging to perform the holographic overlay. Moreover, every small movement and deformation of the kidney will cause an error in the holographic overlay, due to the absence of in-patient tracking. Therefore, it is important to investigate the ability to create a holographic overlay on a more deformable kidney phantom. This is stated in the next chapter. However, the accuracy on a more

deformable phantom has to be investigated in the future, more future perspectives are given in chapter 9.

### 6.5 Conclusion

The goal of this chapter was to validate the home-build AR application. This application is sufficient to be used during clinical validation, stated in chapter 7. Based on the mean measured accuracy of 3.16 millimetre based on 50 points which is within the clinically determined error of 5 millimetre.

## 7 Part 3: Clinical validation

### 7.1 Introduction

In this chapter, the third and last measurement part of this thesis will state the clinical validation of the application. It is important to investigate the usability of this application in a more clinical setting. This will improve the understanding of possible use of this application in the future. This part of the thesis can be performed, due to a sufficient assessed accuracy of the application during the latter parts. The assessed accuracy is within the maximally allowed clinical error. Therefore, the surgical application is sufficient to be used during a clinical validation study. During the other two parts, a couple of important outcomes have emerged which have to be taken into account before we can conclude if the system is sufficient for using for NSS specific. Due to the results of the first part, it is known that the QR-marker dimensions cannot be decreased without a significant decline in accuracy. The results of the second part showed that the system has an accuracy of 3.16 mm, which is less than the maximal clinical error. However, this part was conducted on a rigid 3D printed phantom, which does not simulate the real surgical environment.

Notwithstanding, it is important to know if it would be possible to implement this application in the clinical workflow during NSS. After the first part it appeared to be impossible to use a QR-marker for direct live tracking of the kidney, due to the minimal dimensions of the QR-marker. Therefore, the application should be used only when the kidney is positioned perfectly for the tumour resection. After the registration the kidney should not be moved, or the registration phase should be repeated. However, during surgery, time is limited. Therefore, it is necessary that the surgeon can perform the registration within minutes and accurate when the kidney is positioned perfectly.

Moreover, the feasibility of this application by the surgeons is essential. Without their approval, the application will not be used during surgery. It is necessary to get all surgeons familiar with this application and to get their feedback. They are aware of the important details which evaluate the ability to implement new techniques into the clinical workflow. They are the future users of the device and could have additional requirements to increase the usability of the application. These requirements and the usability should scientifically be evaluated for improving the application.

Therefore, this part of the study is designed to get sight of the usability and future implementation opportunities, based on the opinions of all paediatric oncologic surgeons of the Princess Máxima Centre.

### 7.2 Materials and methods

To simulate the surgical environment, a 3D abdominal surgical phantom was created. This phantom consists of a liver, spleen, spine and abdominal wall made out of silicon (Additive Shore 5 mixed with Shore 15, Silicones and more, Geleen, The Netherlands). These organs are created based on the MRI scans of a real patient and are created with a 1:1 scale.

The deformable kidney phantom is made out of hydrogel. These hydrogels are made of Poly(vinyl alcohol) with a 9% weight to water ratio (Sigma-Aldrich, Merck KGaA, Darmstadt, Germany). The tumour has three freeze cycles, compared with two freeze cycles of the kidney. This difference is applied to create a realistic elastic mimic of the tumour and kidney, the tumour is stiffer due to the additive freeze cycle. [83] The arterial vasculature is 3D printed with 70A Filaflex Ultra-Soft (Recreus Industries S.L., Elda, Spain) to mimic the elasticity of real vessels. These 3D printed vessels are hollow, and printed together with the casting mold of the kidney phantom. An overview of the abdominal phantom is seen in figure 10.

Four kidney phantoms, based on four different patients, are used during this measurement. All kidneys are provided with a tumour. These tumours all have different dimensions. Three phantoms are based on patients with a WT in their right kidney, and the remaining phantom is of a left kidney.

Three out of five paediatric oncologic surgeons were observers in the second part of this study and did not need an introduction on how to use the HoloLens. The other two surgeons are introduced with the HoloLens in the same way as the observers in the second part of the study got their introduction, using the in-build HoloLens application “tips”. All observers are asked to perform the in-build eye calibration. After introducing the HL, the measurements are explained to the observer. Moreover, they get a description of the questionnaire, which they have to fill in afterwards. This questionnaire is added in Appendix 12.2.

The first step of this measurement is the optimal patient mobilisation. The surgeons can use two clamps, which are fixated onto a table with a GoPro arm to create mobility to the clamp. These clamps are used to mimic a Thompson retractor, which is used during nephron-sparing surgery. [73] The clamps can be used to fixate the liver and skin of the abdominal phantom to establish the optimal view on the kidney phantom, this is visualised in figure 10. After the placement of the clamps, the surgeon is asked to mobilize the kidney as he or she would do if he would perform NSS. Subsequently, the “Kidney registration and tracking” application is started.

The observers can choose between the four patients when opening the “Kidney registration and tracking” application. After choosing the patient, they have to select the virtual landmarks they are able to pinpoint onto the phantom, without moving the phantom. Subsequently, the surgeon will start the registration and tracking phase. He will pinpoint the landmarks on the kidney phantom. This is seen in figure 10. If the overlay is sufficient, the kidney phantom will be replaced by the next kidney phantom, until all phantoms are investigated.

Afterwards, the surgeon will receive the questionnaire. This questionnaire consists of three categories. The first category is based on the raw NASA task load index (RTLX). This part is designed to get an insight on how demanded the use of this application is. The RTLX consists of six subjective subscales serving six parts of demand: mental demand, physical demand, temporal demand, performance, effort and frustration. These six questions combined are used to assess the perceived workload of registration task. [84] A low RTLX-score suggest that the application is simple in use. The RTLX-scores will be compared to the RTLX-scores of other medical tasks.

The second category consist of thirteen Likert-scale questions, to determine application specific experiences of the surgeons. These questions are focussing on the user-friendliness, interfaces, clinical usability and future perspective. The outcomes of these questions are used to evaluate the ability of clinical use in the future of our application. The 5-point Likert-scale questions are formulated in a way that a high score represents a positive experience and a low score a negative experience.

The last part is designed with open questions to get the opinion of the surgeons. The questions are focussing on improvements of this application and the future perspectives. The surgeons are asked to give their professional opinion on the shortcomings of the application, and necessary improvements before possible clinical implementation. For an analysis for ability of the clinical implementation the surgeons are asked to note their total amounts of tries needed for a sufficient overlay. The Questionnaire is added in appendix 12.2.



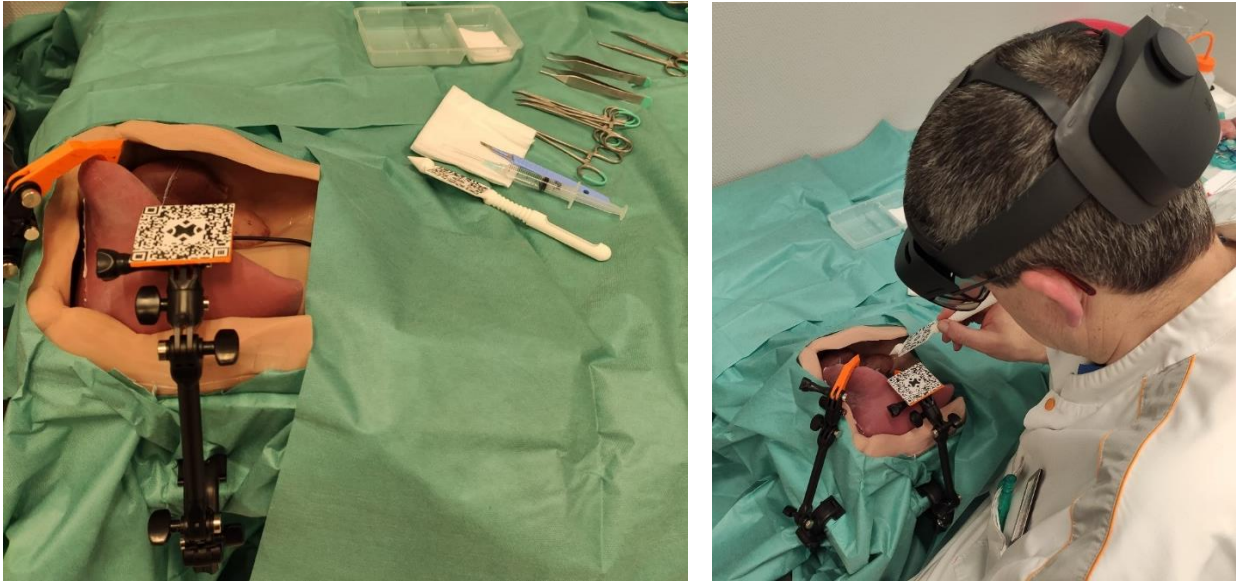


Figure 10, in the left figure the phantom is visualised ready for use. The right part of the image showed one of the surgeons during the registration process.

### 7.3 Results

All surgeons were able to get a sufficient holographic overlay on all four kidney phantoms. One of these overlays can be seen in figure 11. Not all first assessments were sufficient. The mean number of registration attempts per phantom were 2.2, 1.6, 1.2, and 2.

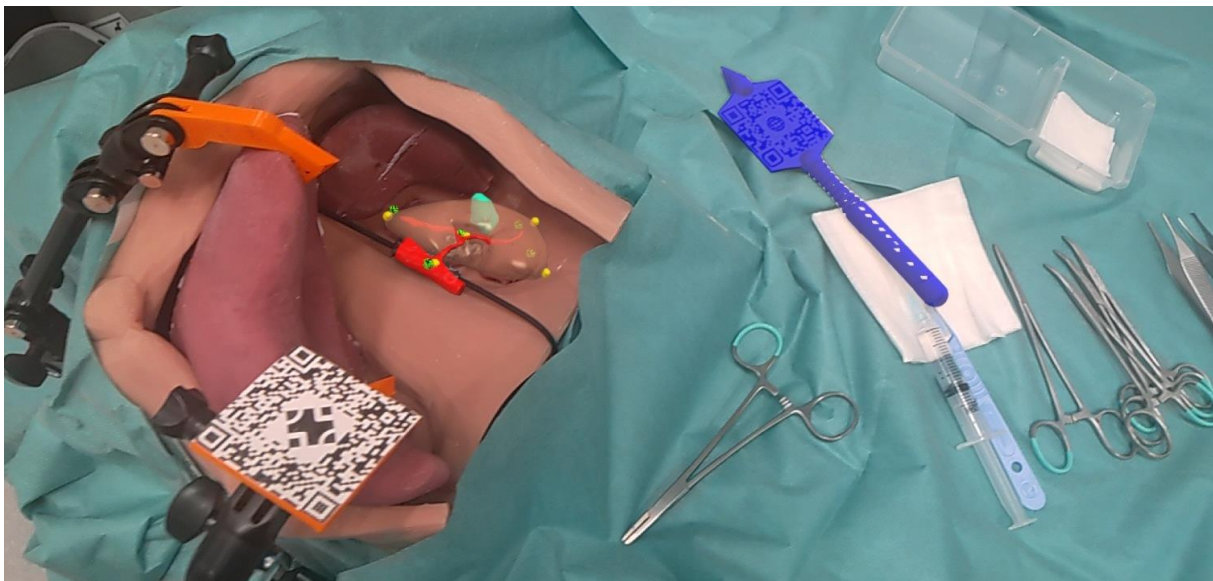


Figure 11, the visualisation after applying the registration method onto the abdominal phantom, visualised through the HoloLens 2.

The RTLX is visualised in figure 12. This figure showed the RTLX score of every surgeon divided per category. The minimal RTLX score is 10.8 and the maximal RTLX score is 59.2, the mean RTLX value of all surgeons is 35.

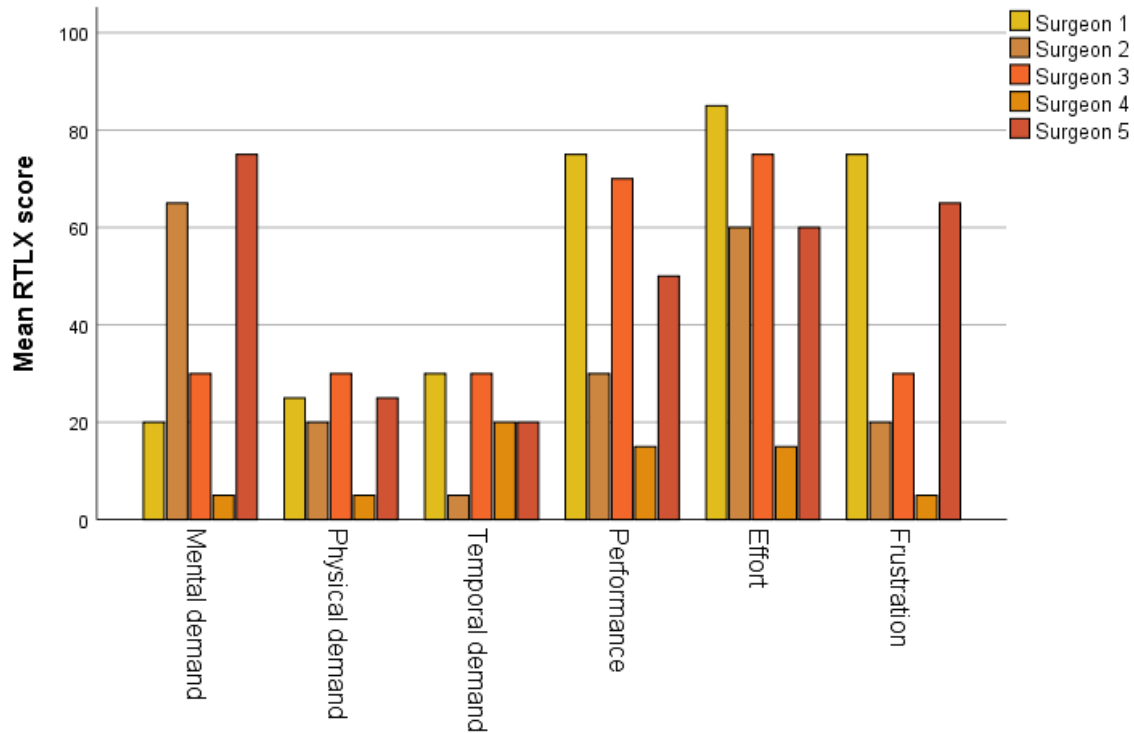


Figure 12, the RTLX-scores per observer per category of the Nasa Raw TLX based on the use of the home-build augmented reality application.

The results of the Likert scale questions are visualised in table 6. Question seven is formulated as: “the virtual overlay did not hinder the view of the surgical site”.

Table 6, the scores of the Likert-scale questions given by the five paediatric surgeons after using the home-build augmented reality application.

Question	Surgeon 1	Surgeon 2	Surgeon 3	Surgeon 4	Surgeon 5	Mean value
1	4	5	4	5	3	4.2
2	3	5	5	5	2	4
3	2	3	3	4	4	3.2
4	4	4	5	5	4	4.4
5	4	2	3	4	4	3.4
6	2	3	3	4	3	3
7	2	1	2	2	2	1.8
8	4	2	4	4	4	3.6
9	3	4	4	4	4	3.8
10	3	4	4	5	3	3.8
11	4	3	5	5	3	4
12	4	4	5	5	2	4
13	4	3	5	5	3	4

Based on the questionnaires, all surgeons experienced a learning curve during this experiment. One of the surgeons answered with, “yes very steep”. The second question was not answered unanimously,

three of the surgeons did not experienced any differences between right or left kidneys, while the two others did experience a difference. Both experienced a worse outcome at right kidneys. One explained this was due to problems to pinpoint the lateral landmark due to the presence of the liver. According to one surgeon there was a difference between small and large tumours as well. She answered that it was a slightly more difficult in a big tumour because the lateral landmark is difficult to recognise due to the lateral location of the large tumour. The other surgeons did not see any difference between small and large tumours.

All surgeons answered they will use this application in the future during NSS. They see an added value for visualising the tumour location and tumour depth. However, they mentioned a couple of improvements. The accuracy could be better, especially for the renal artery and aorta. The overlay of these vessels did not correspond with the vessels in the phantom. Another improvement suggestion was to use other anatomical landmarks.

The surgeons also see other surgical procedures as a future possibility to use the HoloLens and this application, such as, liver metastasectomy, partial long and liver resections, and bone cancer surgery. They see the added value of live 3D visualisation of during these surgical procedures. During these procedures it should be possible to pinpoint predefined anatomical landmarks in the real world.

### 7.4 Discussion

The first part of the questionnaire consists of the raw NASA task load index, this is a widely used method to assess the perceived workload in the healthcare sector. The task load of medical tasks are varying between 9 and 77.35. [85], [86] The augmented reality application used in this study had a minimal value of 10.8 and a maximal value of 59.2. The RTLX-scores difference between the different subcategories. The observers did not experience a physical demand or rushed during this task, visualised with the physical demand and temporal demand. The surgeons rated the effort with a mean value of 59/100 above the average. This subcategory is stated as; “How hard did you have to work to accomplish your level of performance”. Besides the effort subcategory, the performance rated also with an elevated score. This category is stated as; “How successful were you in accomplishing what you were asked to do?”. By combining these two subcategories, it can be concluded that the surgeons had relative high workload to reach a sufficient performance. However, retrospectively the surgeons stated that the elevated scores of these subcategories was mainly caused by the inexperience with the HoloLens 2. Overall, the workload when using the HoloLens 2 application is not more demanding as other medical tasks.

The second part of the questionnaire consists of thirteen Likert-scale questions. The surgeons agreed with most statements. However, the surgeons did not agree with the statement “the virtual overlay did not hinder the view of the surgical site”. The explanation of the surgeons is based on the brightness of the overlay. Due to the relative high brightness the surgeons could not see the phantom. This is detrimental, because the surgeon needs to see the kidney during NSS. This problem could be easily solved by adding a brightness bar, where the surgeon can change the brightness during surgery.

The surgeons experienced an offset of the holographic overlay of the aorta and the renal artery, while the overlay of the kidney was acceptable. This offset can be explained with the deformation ability of the holographic visualisation. The holographic visualisation of the kidney and the vessels are a rigid model. Meaning the ratio between the structures will stay the same, these structures cannot move separately of each other. The kidney phantom consists of an aorta and renal artery, which mimic the elasticity of real vessels. The orientation of the vessels and the kidney in this phantom can differ from

the orientation in the holographic model. Four of the five landmarks are located around the kidney, and one on the bifurcation of the renal artery from the aorta. The four anatomical landmarks located onto the kidney can be located perfectly on the phantom, corresponding with the holographic model. However, when the orientation between the aorta and the kidney is changed, the last anatomical landmark will not have a sufficient correspondence with the virtual landmark. The brute-forcing method calculates the best conforming overlay. This is achieved when the four landmarks located onto the kidney are linked to their corresponding virtual landmarks. The majority of the points will correspond, thus the overlay seems to be sufficient except for the aortic overlay. What will still be in the original orientation relative to the kidney and therefore is not corresponding with the actual aorta.

During the clinical validation all surgeons were able to get a sufficient overlay on all four phantoms. However, they did not achieve this in one try. There were multiple reasons to retry the registration. Firstly, two observers did not have prior experience with this AR application. Therefore, they need to get familiar with this application and the registration method. Secondly, the surgical pointer was moved too soon after the voice command “fix point”. Resulting in a wrongly placed landmark. Another reason was the use of too little landmarks, which resulted in a rotated virtual overlay. Sometimes, the surgeons were convinced they could improve the placement of the landmarks.

To conclude, the surgeons were positive about this application. They definitely see an added for this application during NSS and other surgical procedures. However, some improvements have to be made before this application could be implemented into the clinical workflow.

### 7.5 Conclusion

During these measurements the application is tested on the ability to perform an accurate overlay onto a phantom with the mobility of a real kidney, and in an environment similar to an abdomen. The surgeons are enthusiastic about the possibilities of the HoloLens during nephron-sparing surgery. The registration method is usable during NSS. Therefore, this application has the ability to become an added value during NSS. However, some improvements have to be made first.

## 8 Overall discussion

The goal of this study was to design and evaluate a new AR driven surgical navigation system specifically for the application during NSS. The intraoperative 3D visualisation of the tumour, kidney and critical vessels could help with the determination of the resection margins. For this purpose, an application for the HoloLens 2 was build.

The accuracy of the use of QR-markers as a tracking method was successfully measured during the first experiments. The dimension of the best working QR-marker was 5x5 cm, which therefore is used in the further measurements. However, this dimension is not suitable for in-patient tracking. A trackable surface of 25cm<sup>2</sup> is too big compared to the size of a paediatric kidney. A healthy paediatric kidney size can be calculated with the following equation:  $y = 0.379x + 6.65$ . [87] For this equation y reference to the length of the kidney in centimetre, and x to the age of the patient in years. This equation is usable for children between the age of five and thirteen years old. This means that the length of a kidney of a five year old corresponded with approximately 8.5 centimetres. Consequently, the QR-marker will cover more than half of the kidney size. This will hinder the surgical site. Therefore, the use of QR-marker tracking is not suitable for in patient tracking during NSS.

The use of the QR-marker tracking on the surgical pointer is suitable. For the surgical pointer the dimensions of the QR-marker are less critical. It is important that the surgeon is able to move the surgical pointer without losing the tracking. The HoloLens should registrate the QR-marker in different angles as well. This will optimise the degrees of freedom for the surgeon to move and rotate the surgical pointer. This optimised the ability to place the anatomical landmarks as close to the virtual landmarks as possible. Small adjustments should be made to optimise the surgical pointer. The tooltip is too sharp for pinpointing onto human tissue, and the edges are too sharp, which increases the risk on tissue damage. The use of the QR-marker tracked surgical pointer for a landmark based registration method seems to be sufficient.

47 of the 50 marked points set in chapter 6 are within the clinical maximal error. This suggests that the registration method is working properly. During the analysis of these marked points, it was expected that there is probably an overestimate of the measured error. The measured error based solely on the registration method is smaller than the measured 3.16 mm. However, this overestimation could have a negative influence during NSS. The overestimation is based on the interpretation of the holographic visualisation by the observers. This interpretation also has a huge influence on the accuracy of the tumour resection. The surgeons decide the location of the resection margins. They will base this decision on their interpretation of the holographic visualisation. If the surgeon interprets the location of the tumour more caudal, due to the tumour location, a shift of the hologram could arise. A shift in this interpretation, could lead to a wrongly determined resection due to a poor understanding of the location of the tumour. A wrongly determined resections leads to a negative influence on the clinical patient outcome. Therefore, it is important to investigate the influence of the interpretation of the surgeons on the accuracy of this HoloLens application, recommendations on future studies can be found in chapter 9.

The surgeons were content with the current HoloLens application for use during NSS. They see a role for this application and the HoloLens in the OR in the future. The task load of this application is sufficient, and the application is user-friendly, according to the surgeons based on the NASA raw task load index. The NASA RTLX is broadly adopted as a workload assessment. [88] However, it is not specifically developed to assess the workload during surgery. Therefore, we cannot conclude that the workload of this application is sufficient during surgery. Wilson et al (2011) created a task load index specifically for surgery. [89] This questionnaire could have an additional value for the analysis of

the workload of this surgical navigation application. Wilson et al. implemented multiple factors of the OR which are not implemented in the RTLX, and which are not simulated during this measurement. For instance, during NSS the surgeon has a higher stress level, this will increase the RTLX score of this application. Moreover, during this measurement the surgeons did not experienced a high temporal demand. However, during NSS the surgeon has 30 minutes between the ligation of the vessels and the total tumour resection, this will probably increase the experienced temporal demand. To analyse the real RTLX for this application, it should be measured during NSS. Another factor what will influence the RTLX score is the learning curve of this application. The surgeons all experienced a steep learning curve. However, they had a maximal experience of 1.5 hours with this HL2 application. We suggested that the learning curve is not maximised with this 1.5 hour. When the surgeons increase their experience with the HoloLens 2, the RTLX will probably decrease. Based on the clinical validation, the surgeons had some easily adjustable and some more advanced adjustable improvement suggestions for the application. These suggestions will be further explained in chapter 9.1 and 9.2. If these recommended adjustments are successfully implemented, next steps could be made towards clinical implementation. These steps are explained in chapter 9.3.

There are multiple limitations in the measurements explained in chapter 5, 6 and 7. It would been better if the measurements of chapter 5 had taken place at the same location as the measurements of chapter six and seven. The results of these chapters would be more comparable if all measurements would have the same measurement environment, and thus increase the accuracy of the analysis. Moreover, the lighting conditions in this room improve the registration and tracking ability of the HoloLens 2. This could increase the accuracy of the 4.5x4.5 cm QR-marker, which could differ the choice of the QR-marker dimensions.



## 9 Overall recommendations / future perspectives

Multiple adjustments and improvements have to be made before this application is suitable for clinical implementation. In this chapter these adjustments and improvements will be divided in three subcategories, easily technical adjustable improvements, more complicated technical improvements and next steps to clinical implementation.

### 9.1 Short-term technical improvements

The first easily adjustable improvement is the exclusion of the vessels in the 3D holographic model. This is already been discussed in section 7.4. When excluding the vessels an anatomic landmark is deleted as well. A decreasing in the number of used landmarks will result in a less stable registration method. A wrongly placed anatomic landmark by the surgeon, will have a greater effect on the registration if a lower number of landmarks is used. However, based on step 2 of section 4.4, increasing the number of landmarks will increase the required computer power with a factor 2. This will result in a longer registration time, and thus a longer surgery time. Therefore, it is not desirable to use more intraoperative anatomical landmarks. The recommendation of this improvement is to exclude the vessels, and to replace the landmark of the aorta to the medial side of the kidney.

The second easily adjustable improvement is the implementation of a brightness bar, also already discussed in section 7.4. As stated in section 7.4, the holographic overlay hindered the view of the surgical site. Therefore, it is important to implement a method by which the surgeon can adjust the brightness of the hologram. During the identification of the tumour by the surgeon through the HoloLens, it is desirable to have a bright hologram. However, during the tumour resection it is desirable for the surgeon to see the anatomical structures as well. Therefore, a bar by which the surgeon could adjust the brightness of the hologram is desirable.

### 9.2 Long-term technical improvements

Some technical improvements are not implementable on short-term. These improvements would have serious added value for the application. During this section, two main complicated technical improvements will be mentioned. Firstly, a new tracking method should be investigated. The QR-marker recognition is working properly. However, the dimensions of the QR-marker are too large to fixate it onto the kidney, thus the kidney is not being tracked when it moves. Therefore, the registration has to be reperformed after movement. It would be an added value if the kidney can be tracked and a second registration is not necessary.

The first recommendation is to look into is the use of infrared (IR) tracking. IR markers are already been used in multiple surgical procedures, such as neurosurgery. Multiple medical technology companies have IR tracking devices, such as Brainlab and Stryker. [60], [90] However, little is known about the combination of IR tracking with the HL 2. Nevertheless, the use of IR markers could be useful for the use of intraoperative tracking, if the markers are detachable onto the kidney. The IR markers have a smaller surface than the QR-markers, thus are not a burden for the surgeon during NSS. However other tracking method with smaller markers, or markerless tracking methods could also be investigated.

Another long-term technical improvement which should be investigated is a method to include organ deformation into this application. The used methods in this application are rigid. However most of the human body do not respond conform to rigid models. Rigid models works properly until the target organ is deformed during surgery. [91] The kidney is a soft-tissue organ, therefore it is prone to

intraoperative organ deformations. These deformations are caused by changes in pressure due to patient's position, or clamping of the renal vessels, or due to organ manipulation or resection by the surgeon. [92] These factors result in a decreased level of matching.

Multiple research groups are conducting research to a non-rigid registration method for augmented reality for different medical applications. Morelli et al. (2020) [92] are developing a non-rigid registration method for augmented reality in nephron-sparing surgery. Their goal is to develop a deformable registration strategy for intraoperative use during nephrectomy. The used deformable registration method is based on the idea to represent the deformation of the underlying mesh. Golse et al. (2020) [93] developed a deformable model based on the elasticity of the liver. This model converts the internal elastic forces into a deformation based on a stiffness matrix. Another possible implementation of a non-rigid augmented reality method is the combination with optical 3D markers. Adagolodjo et al. (2018) [94] used this method during open liver surgery. All these different non-rigid methods provided promising results, however, further development is necessary. [92]–[94]

### 9.3 Next steps to clinical implementation

Prior to the use of this application intraoperative, several measurements have to be performed, to assure the safety of the patients and the added value of this application. The first measurement is the measurement of the accuracy of this application within an environment similar to NSS. During this thesis the accuracy has only be measured on a rigid phantom. It is important to know if the system is user dependent during measurements on a non-rigid phantom.

One measurement to test the user dependency is to conduct NSS on the abdominal phantom used in chapter 7. Different surgeons can be asked to perform NSS onto the phantom with and without the use of the HoloLens 2. The results of these NSS can be analysed based on the presence of positive surgical margins. However, these margins are not detectable using the kidney phantoms used in this study. To simulate the clinical workflow, it would be desirable to put the phantoms in the MRI scanner. However, the phantoms used in this study are made out of hydrogels, the tumours as well. It is desirable to adjust the kidney phantoms for a better visualisation in the MRI scanner. The preoperative and postoperative MRI could be compared to investigate the surgical margins. However, the deformation of the phantoms after resection should be taken into account.

If these measurements have a positive outcome, this experiment could be repeated with porcine kidney. This represents the human kidney probably better than the kidney phantoms. These porcine kidneys have to be injected with a substance that mimics a tumour. The size of the porcine kidney has to be taken into account, it has to be corresponding to the size of a paediatric kidney to experience the perioperative settings as well as possible.

The next step towards clinical implementation would be the use of the HoloLens application in the OR. A clinical study could be started in combination with the pathology department. It is recommended to start with a study using the HoloLens after radical nephrectomy. The surgeon could use the HoloLens after resection, to visualise the kidney, tumour and critical vessels. With the information provided by the HoloLens the surgeon could mark the resection margin, which would be used if NSS was performed. In collaboration with the pathology department the resection margins could be assessed compared with the histology. This study design will not increase the surgical risks of the patients. A clinical study will provide information on the usability of this application in the OR, and on the accuracy of the application based on human tissue



## 10 Overall conclusion

To conclude, we have developed an AR surgical navigation system based on a point-based registration method and a QR-marker tracking method for intraoperative use to guide surgeons during NSS of WT patients. These methods have a mean combined accuracy of 3.16 mm and a standard deviation of 1.82 mm, which meets the maximally clinically error of 5 mm set by the surgical team of the Princess Máxima Centre. Therefore, this HoloLens 2 surgical navigation system is evaluated to be technically sufficient for this surgical application. The tracking method should be reconsidered, due to the disability to use the tracking method in-patient, due to the minimal dimensions of the QR-marker. On the contrary, the registration method is working properly.

The surgical navigation system is user-friendly and of added value in visualisation of the kidney, tumour and critical vessels, according to all five paediatric surgeons of the Princess Máxima Centre. Some improvements should be made before this application could be used intraoperatively.

## 11 References

- [1] N. Breslow, A. Olshan, *et al.*, "Epidemiology of Wilms tumor," *Med. Pediatr. Oncol.*, vol. 21, no. 3, pp. 172–181, 1993, doi: 10.1002/mpo.2950210305.
- [2] B. Erginel, "Wilms Tumor and Its Management in a Surgical Aspect," in *Wilm Tumor*, Codon Publications, 2016, pp. 43–59.
- [3] M. E. Cunningham, T. D. Klug, *et al.*, "Global Disparities in Wilms Tumor," *J. Surg. Res.*, vol. 247, pp. 34–51, Mar. 2020, doi: 10.1016/j.jss.2019.10.044.
- [4] M. Malogolowkin, F. Spreafico, *et al.*, "Incidence and Outcomes of Patients With Late Recurrence of Wilms' Tumor," *Pediatr Blood*, vol. 60, pp. 1612–1615, 2013, doi: 10.1002/pbc.24604.
- [5] C. Wei, C. Li, *et al.*, "Performance Characteristics of Transrectal Shear Wave Elastography Imaging in the Evaluation of Clinically Localized Prostate Cancer: A Prospective Study," *J. Urol.*, vol. 200, no. 3, pp. 549–558, Sep. 2018, doi: 10.1016/J.JURO.2018.03.116.
- [6] E. Szychot, J. Apps, *et al.*, "Wilms' tumor: biology, diagnosis and treatment.," *Transl Pediatr*, vol. 3, no. 1, pp. 12–24, 2014, doi: 10.3978/j.issn.2224-4336.2014.01.09.
- [7] H. Chen, S. Yang, *et al.*, "Effectiveness of Nephron Sparing Surgery and Radical Nephrectomy in the Management of Unilateral Wilms Tumor: A Meta-Analysis," *Front. Oncol.*, vol. 10, p. 1248, Sep. 2020, doi: 10.3389/fonc.2020.01248.
- [8] G. Khanna, A. Naranjo, *et al.*, "Detection of preoperative Wilms tumor rupture with CT: A report from the Children's Oncology Group," *Radiology*, vol. 266, no. 2, pp. 610–617, Feb. 2013, doi: 10.1148/radiol.12120670.
- [9] J. H. Aldrink, N. G. Cost, *et al.*, "Technical considerations for nephron-sparing surgery in children: what is needed to preserve renal units?," *J. Surg. Res.*, vol. 232, pp. 614–620, Dec. 2018, doi: 10.1016/j.jss.2018.07.022.
- [10] M. Harel, J. H. Makari, *et al.*, "Oncology: The role of partial nephrectomy in wilms tumor," *Curr. Urol. Rep.*, vol. 14, no. 4, pp. 350–358, Aug. 2013, doi: 10.1007/s11934-013-0330-0.
- [11] J. C. H. Wilde, D. C. Aronson, *et al.*, "Nephron sparing surgery (NSS) for unilateral wilms tumor (UWT): The SIOP 2001 experience," *Pediatr. Blood Cancer*, vol. 61, no. 12, pp. 2175–2179, Dec. 2014, doi: 10.1002/pbc.25185.
- [12] N. Wake, T. Rude, *et al.*, "3D printed renal cancer models derived from MRI data: application in pre-surgical planning," *Abdom. Radiol.*, vol. 42, no. 5, pp. 1501–1509, May 2017, doi: 10.1007/s00261-016-1022-2.
- [13] M. Hegarty, M. Keehner, *et al.*, "The Role of Spatial Cognition in Medicine: Applications for Selecting and Training Professionals.," *Applied spatial cognition: From research to cognitive technology*. Lawrence Erlbaum Associates Publishers, Mahwah, NJ, US, pp. 285–315, 2007.
- [14] H. K. Wu, S. W. Y. Lee, *et al.*, "Current status, opportunities and challenges of augmented reality in education," *Comput. Educ.*, vol. 62, pp. 41–49, Mar. 2013, doi: 10.1016/j.compedu.2012.10.024.
- [15] G. Kipper, "What Is Augmented Reality?," in *Augmented Reality*, Elsevier, 2013, pp. 1–27.
- [16] M. Eckert, J. S. Volmerg, *et al.*, "Augmented reality in medicine: Systematic and bibliographic review," *J. Med. Internet Res.*, vol. 21, no. 4, Apr. 2019, doi: 10.2196/10967.

- [17] M. N. Rivera and D. A. Haber, "Wilms' tumour: Connecting tumorigenesis and organ development in the kidney," *Nature Reviews Cancer*, vol. 5, no. 9. Nature Publishing Group, pp. 699–712, Sep. 19, 2005, doi: 10.1038/nrc1696.
- [18] S. A. Narod, M. M. Hawkins, *et al.*, "Congenital anomalies and childhood cancer in Great Britain," *Am. J. Hum. Genet.*, vol. 60, no. 3, pp. 474–485, Mar. 1997, Accessed: Dec. 04, 2020. [Online]. Available: /pmc/articles/PMC1712528/?report=abstract.
- [19] K. Sanpakit, J. Triwatanawong, *et al.*, "Long-term Outcome in Pediatric Renal Tumor Survivors: Experience of a Single Center," 2013. Accessed: Feb. 05, 2021. [Online]. Available: www.jpho-online.com.
- [20] R. H. Scott, C. A. Stiller, *et al.*, "Syndromes and constitutional chromosomal abnormalities associated with Wilms tumour," *Journal of Medical Genetics*, vol. 43, no. 9. BMJ Publishing Group, pp. 705–715, Sep. 2006, doi: 10.1136/jmg.2006.041723.
- [21] J. M. Gleason, A. J. Lorenzo, *et al.*, "Innovations in the management of Wilms' tumor," *Therapeutic Advances in Urology*, vol. 6, no. 4. SAGE Publications, pp. 165–176, 2014, doi: 10.1177/1756287214528023.
- [22] G. Diniz, "Histopathological and Molecular Characteristics of Wilms Tumor," in *Wilms Tumor*, Codon Publications, 2016, pp. 31–41.
- [23] M. M. Van Den Heuvel-Eibrink, J. A. Hol, *et al.*, "Position Paper: Rationale for the treatment of Wilms tumour in the UMBRELLA SIOP-RTSG 2016 protocol," *Nat. Rev. Urol.*, vol. 14, no. 12, pp. 743–752, Dec. 2017, doi: 10.1038/nrurol.2017.163.
- [24] C. Mitchell, K. Pritchard-Jones, *et al.*, "Immediate nephrectomy versus preoperative chemotherapy in the management of non-metastatic Wilms' tumour: Results of a randomised trial (UKW3) by the UK Children's Cancer Study Group," *Eur. J. Cancer*, vol. 42, no. 15, pp. 2554–2562, Oct. 2006, doi: 10.1016/j.ejca.2006.05.026.
- [25] P. Ehrlich, Y. Y. Chi, *et al.*, "Results of the First Prospective Multi-institutional Treatment Study in Children with Bilateral Wilms Tumor (AREN0534): A Report from the Children's Oncology Group," *Ann. Surg.*, vol. 266, no. 3, pp. 470–478, Sep. 2017, doi: 10.1097/SLA.0000000000002356.
- [26] D. C. Aronson, A. Slaar, *et al.*, "Long-Term Outcome of Bilateral Wilms Tumors (BWT)," *Pediatr Blood Cancer*, vol. 56, pp. 1110–1113, 2011, doi: 10.1002/pbc.22881.
- [27] M. M. Van Den Heuvel-Eibrink, J. A. Hol, *et al.*, "Position Paper: Rationale for the treatment of Wilms tumour in the UMBRELLA SIOP-RTSG 2016 protocol," *Nat. Rev. Urol.*, vol. 14, no. 12, pp. 743–752, Dec. 2017, doi: 10.1038/nrurol.2017.163.
- [28] R. B. Nerli, V. C. Pujar, *et al.*, "Nephron Sparing Surgery for Unilateral Non-Syndromic Wilms Tumor," *Indian J. Surg. Oncol.*, vol. 5, no. 1, pp. 11–16, Mar. 2014, doi: 10.1007/s13193-013-0284-1.
- [29] C. G. M. Moorman-Voestermans, D. C. Aronson, *et al.*, "Is partial nephrectomy appropriate treatment for unilateral Wilms' tumor?," in *Journal of Pediatric Surgery*, Feb. 1998, vol. 33, no. 2, pp. 165–170, doi: 10.1016/S0022-3468(98)90425-0.
- [30] R. N. W. Vanden Berg, E. N. Bierman, *et al.*, "Nephron-sparing surgery for Wilms tumor: A systematic review," *Urologic Oncology: Seminars and Original Investigations*, vol. 34, no. 1. Elsevier Inc., pp. 24–32, Jan. 01, 2016, doi: 10.1016/j.urolonc.2015.07.003.
- [31] M. K. Richards, A. B. Goldin, *et al.*, "Partial Nephrectomy for Nephroblastoma: A National

- Cancer Data Base Review," *Am. Surg.*, vol. 84, no. 3, pp. 338–343, Mar. 2018, doi: 10.1177/000313481808400315.
- [32] N. G. Cost, J. D. Lubahn, *et al.*, "Pathological review of wilms tumor nephrectomy specimens and potential implications for nephron sparing surgery in wilms tumor," *J. Urol.*, vol. 188, no. 4 SUPPL., pp. 1506–1510, 2012, doi: 10.1016/j.juro.2012.02.025.
- [33] B. Erginel, "Wilms Tumor and Its Management in a Surgical Aspect," in *Wilms Tumor*, Brisbane: Codon Publications, 2016, pp. 43–59.
- [34] Brilliantino C, E. Rossi, *et al.*, "Translational Medicine Current Role of Imaging in the Management of Children with Wilms Tumor According to the New UMBRELLA Protocol," *Transl Med*, vol. 8, no. 2, p. 206, 2019.
- [35] T. Klatte, V. Ficarra, *et al.*, "A Literature Review of Renal Surgical Anatomy and Surgical Strategies for Partial Nephrectomy," *European Urology*, vol. 68, no. 6. Elsevier B.V., pp. 980–992, Dec. 2015, doi: 10.1016/j.eururo.2015.04.010.
- [36] S. E. Servaes, F. A. Hoffer, *et al.*, "Imaging of Wilms tumor: an update," *Pediatric Radiology*, vol. 49, no. 11. Springer Verlag, pp. 1441–1452, Oct. 01, 2019, doi: 10.1007/s00247-019-04423-3.
- [37] M. Fitski, J. W. Meulstee, *et al.*, "MRI-Based 3-Dimensional Visualization Workflow for the Preoperative Planning of Nephron-Sparing Surgery in Wilms' Tumor Surgery: A Pilot Study," *J. Healthc. Eng.*, vol. 2020, 2020, doi: 10.1155/2020/8899049.
- [38] A. Murphy and A. Davidoff, "Bilateral Wilms Tumor: A Surgical Perspective," *Children*, vol. 5, no. 10, p. 134, Sep. 2018, doi: 10.3390/children5100134.
- [39] H. R. Spiegl, A. J. Murphy, *et al.*, "Complications Following Nephron-Sparing Surgery for Wilms Tumor," *J. Pediatr. Surg.*, vol. 55, no. 1, pp. 126–129, Jan. 2020, doi: 10.1016/j.jpedsurg.2019.09.066.
- [40] F. Alharbi, C. Chahwan, *et al.*, "Intraoperative ultrasound control of surgical margins during partial nephrectomy," *Urol. Ann.*, vol. 8, no. 4, pp. 430–433, Oct. 2016, doi: 10.4103/0974-7796.192107.
- [41] C. Lupulescu and Z. Sun, "A Systematic Review of the Clinical Value and Applications of Three-Dimensional Printing in Renal Surgery," *J. Clin. Med.*, vol. 8, no. 7, p. 990, Jul. 2019, doi: 10.3390/jcm8070990.
- [42] K. Milford, K. DeCotiis, *et al.*, "Wilms tumor: A review of current surgical controversies," *Translational Andrology and Urology*, vol. 9, no. 5. AME Publishing Company, pp. 2382–2392, Oct. 01, 2020, doi: 10.21037/tau.2020.03.03.
- [43] H. Abdelhafeez and S. Abib, "IPSO Practice Guidelines on Wilms Tumor - IPSO," *International society of Paediatric surgical oncology*, Jun. 07, 2020. <https://ipso-online.org/guidelines-wt/> (accessed Dec. 15, 2020).
- [44] K. Kieran and A. M. Davidoff, "Nephron-sparing surgery for bilateral Wilms tumor," *Pediatric Surgery International*, vol. 31, no. 3. Springer Verlag, pp. 229–236, Feb. 19, 2015, doi: 10.1007/s00383-015-3668-1.
- [45] Y. Soliman, A. H. Feibus, *et al.*, "3D Printing and Its Urologic Applications," *Rev. Urol.*, vol. 17, no. 1, p. 20, 2015.
- [46] N. Wake, J. S. Wysock, *et al.*, "'Pin the Tumor on the Kidney:' An Evaluation of How Surgeons Translate CT and MRI Data to 3D Models," *Urology*, vol. 131, pp. 255–261, Sep. 2019, doi:

- 10.1016/j.urology.2019.06.016.
- [47] L. M. Wellens, J. Meulstee, *et al.*, “Comparison of 3-Dimensional and Augmented Reality Kidney Models With Conventional Imaging Data in the Preoperative Assessment of Children With Wilms Tumors,” *JAMA Netw. open*, vol. 2, no. 4, p. e192633, Apr. 2019, doi: 10.1001/jamanetworkopen.2019.2633.
  - [48] D. R. Berryman, “Augmented Reality: A Review,” *Medical Reference Services Quarterly*, vol. 31, no. 2, pp. 212–218, Apr. 2012, doi: 10.1080/02763869.2012.670604.
  - [49] A. M. Clark and M. T. G. Clark, “PokémonGo and Research: Qualitative, Mixed Methods Research, and the Supercomplexity of Interventions,” doi: 10.1177/1609406916667765.
  - [50] B. Fida, F. Cutolo, *et al.*, “Augmented reality in open surgery,” *Updates in Surgery*, vol. 70, no. 3, Springer-Verlag Italia s.r.l., pp. 389–400, Sep. 01, 2018, doi: 10.1007/s13304-018-0567-8.
  - [51] P. Vávra, J. Roman, *et al.*, “Recent Development of Augmented Reality in Surgery: A Review,” *J. Healthc. Eng.*, vol. 2017, p. 9, 2017, doi: 10.1155/2017/4574172.
  - [52] J. W. Yoon, R. E. Chen, *et al.*, “Augmented reality for the surgeon: Systematic review,” *Int. J. Med. Robot. Comput. Assist. Surg.*, vol. 14, no. 4, p. e1914, Aug. 2018, doi: 10.1002/rcs.1914.
  - [53] F. Cutolo, N. Cattari, *et al.*, “Optical See-Through Head-Mounted Displays With Short Focal Distance: Conditions for Mitigating Parallax-Related Registration Error,” *Front. Robot. AI*, vol. 7, p. 572001, Dec. 2020, doi: 10.3389/frobt.2020.572001.
  - [54] D. G. Armstrong, T. M. Rankin, *et al.*, “A heads-up display for diabetic limb salvage surgery: A view through the google looking glass,” *J. Diabetes Sci. Technol.*, vol. 8, no. 5, pp. 951–956, Jan. 2014, doi: 10.1177/1932296814535561.
  - [55] T. Bussink, “Augmented reality in cranio-maxillofacial surgery,” University of Twente, 2020.
  - [56] M. Zeller, E. Miller, *et al.*, “HoloLens (1st gen) hardware | Microsoft Docs,” *Microsoft*, 2019. <https://docs.microsoft.com/en-us/hololens/hololens1-hardware> (accessed Feb. 01, 2021).
  - [57] L. Qian, A. Barthel, *et al.*, “Comparison of optical see-through head-mounted displays for surgical interventions with object-anchored 2D-display,” *Int. J. Comput. Assist. Radiol. Surg.*, vol. 12, no. 6, pp. 901–910, Jun. 2017, doi: 10.1007/s11548-017-1564-y.
  - [58] R. Rahman, M. E. Wood, *et al.*, “Head-Mounted Display Use in Surgery: A Systematic Review,” *Surgical Innovation*, vol. 27, no. 1, SAGE Publications Inc., pp. 88–100, Feb. 01, 2020, doi: 10.1177/1553350619871787.
  - [59] E. Miller and S. Kerawala, “HoloLens 2 editions | Microsoft Docs,” *Microsoft*, 2021. <https://docs.microsoft.com/en-us/hololens/hololens2-options?tabs=device> (accessed Jan. 13, 2021).
  - [60] C. M. Andrews, A. B. Henry, *et al.*, “Registration Techniques for Clinical Applications of Three-Dimensional,” *IEEE J. Transl. Eng. Heal. Med.*, vol. 9, 2020, doi: 10.1109/JTEHM.2020.3045642.
  - [61] J. Kosciuszka, “HoloLens 2 vs HoloLens 1: what’s new?,” *4Experience*, 2020. <https://4experience.co/hololens-2-vs-hololens-1-whats-new/> (accessed Feb. 01, 2021).
  - [62] J. W. Yoon, R. E. Chen, *et al.*, “Augmented reality for the surgeon: Systematic review,” *Int. J. Med. Robot. Comput. Assist. Surg.*, vol. 14, no. 4, p. e1914, Aug. 2018, doi: 10.1002/rcs.1914.
  - [63] J. W. Meulstee, J. Nijsink, *et al.*, “Toward Holographic-Guided Surgery,” *Surg. Innov.*, vol. 26, no. 1, pp. 86–94, Feb. 2019, doi: 10.1177/1553350618799552.

- [64] T. Kuzhagaliyev, M. Janatka, *et al.*, “Augmented reality needle ablation guidance tool for irreversible electroporation in the pancreas,” in *Medical Imaging 2018: Image-Guided Procedures, Robotic Interventions, and Modeling*, Mar. 2018, vol. 10576, p. 30, doi: 10.1117/12.2293671.
- [65] I. Kuhlemann, M. Kleemann, *et al.*, “Towards X-ray free endovascular interventions - Using HoloLens for on-line holographic visualisation,” in *Healthcare Technology Letters*, 2017, vol. 4, no. 5, pp. 184–187, doi: 10.1049/htl.2017.0061.
- [66] F. Incekara, M. Smits, *et al.*, “Clinical Feasibility of a Wearable Mixed-Reality Device in Neurosurgery,” *World Neurosurg.*, vol. 118, pp. e422–e427, Oct. 2018, doi: 10.1016/j.wneu.2018.06.208.
- [67] T. P. C. Van Doormaal, J. A. M. Van Doormaal, *et al.*, “Clinical accuracy of holographic navigation using point-based registration on augmented-reality glasses,” *Oper. Neurosurg.*, vol. 17, no. 6, pp. 588–593, 2019, doi: 10.1093/ons/opz094.
- [68] T. Peters, C. Linte, *et al.*, “Mixed and Augmented Reality in Medicine,” *CRC Press, Taylor & Francis Group*, 2019.  
[https://books.google.nl/books?id=e\\_Z1DwAAQBAJ&pg=PA33&lpg=PA33&dq=Landmark-based+registration+Artificial+reality+medical&source=bl&ots=\\_sukOfymDx&sig=ACfU3U2tt3rqYHIEytYRVmiYk5fRRVRf1w&hl=en&sa=X&ved=2ahUKewjzt47K3tPpAhXKDOWKHcrZBUoQ6AEwAXoECAsQAQ#v=onepage&q=Landmark-based registration Artificial reality medical&f=false](https://books.google.nl/books?id=e_Z1DwAAQBAJ&pg=PA33&lpg=PA33&dq=Landmark-based+registration+Artificial+reality+medical&source=bl&ots=_sukOfymDx&sig=ACfU3U2tt3rqYHIEytYRVmiYk5fRRVRf1w&hl=en&sa=X&ved=2ahUKewjzt47K3tPpAhXKDOWKHcrZBUoQ6AEwAXoECAsQAQ#v=onepage&q=Landmark-based registration Artificial reality medical&f=false) (accessed Feb. 05, 2021).
- [69] K. Shree Mudgal and N. Das, “Augmented Reality in Kidney Cancer,” in *Kidney Cancer [Working Title]*, IntechOpen, 2018.
- [70] A. Seginer, “Rigid-body point-based registration: The distribution of the target registration error when the fiducial registration errors are given,” *Med. Image Anal.*, vol. 15, no. 4, pp. 397–413, Aug. 2011, doi: 10.1016/j.media.2011.01.001.
- [71] T. Frantz, B. Jansen, *et al.*, “Augmenting microsoft’s HoloLens with vuforia tracking for neuronavigation,” *Healthc. Technol. Lett.*, vol. 5, no. 5, pp. 221–225, 2018, doi: 10.1049/htl.2018.5079.
- [72] T. Bien, M. Li, *et al.*, “Electromagnetic tracking system with reduced distortion using quadratic excitation,” *Int. J. Comput. Assist. Radiol. Surg.*, vol. 9, no. 2, pp. 323–332, 2014, doi: 10.1007/s11548-013-0925-4.
- [73] Thompson surgical, “Thompson Surgical Instruments,” *Thompson surgical*, 2021.  
<https://thompsonsurgical.com/> (accessed Jan. 18, 2021).
- [74] M. Maylyan, B. Ho, *et al.*, “Anchor your holograms to the real world in Dynamics 365 Guides - Dynamics 365 Mixed Reality | Microsoft Docs,” *Microsoft*, 2020.  
<https://docs.microsoft.com/en-us/dynamics365/mixed-reality/guides/anchor> (accessed Jan. 18, 2021).
- [75] B. Paláncz, P. Zaletnyik, *et al.*, “Extension of the ABC-Procrustes algorithm for 3D affine coordinate transformation,” *Earth, Planets Sp.*, vol. 62, no. 11, pp. 857–862, Nov. 2010, doi: 10.5047/eps.2010.10.004.
- [76] G. Luijten, “Holographic augmented reality in deep inferior epigastric perforator flap breast reconstructions,” *University of Twente*, 2020.
- [77] Vuforia Developer Library, “Device Tracking Overview | VuforiaLibrary,” *PTC Inc.*, 2020.  
<https://library.vuforia.com/features/environments/device-tracker-overview.html> (accessed



- Jan. 22, 2021).
- [78] Technical committee ISO, "ISO 5725-4:2020(en), Accuracy (trueness and precision) of measurement methods and results — Part 4: Basic methods for the determination of the trueness of a standard measurement method," Mar. 2020.  
<https://www.iso.org/obp/ui/#iso:std:iso:5725:-4:ed-2:v1:en> (accessed Feb. 03, 2021).
  - [79] D. Brown and L. Oreilly, "HoloLens environment considerations | Microsoft Docs," *Microsoft*, 2019. <https://docs.microsoft.com/en-us/hololens/hololens-environment-considerations> (accessed Jan. 27, 2021).
  - [80] A. Putta, "Implementation of Augmented Reality applications to recognize Automotive Vehicle using Microsoft HoloLens Performance comparison of Vuforia 3-D recognition and QR-code recognition Microsoft HoloLens applications," Blekinge Institute of Technology, 2019.
  - [81] T. Huysmans and J. Molenbroek, "DINED anthropometric database," *TU Delft*, 2000.  
<https://dined.io.tudelft.nl/en/database/tool> (accessed Jan. 27, 2021).
  - [82] Laerd statistics, "One-way ANOVA - How to report the significance results, homogeneity of variance and running post-hoc tests | Laerd Statistics," *Lund Research Ltd*, 2018.  
<https://statistics.laerd.com/statistical-guides/one-way-anova-statistical-guide-4.php> (accessed Jan. 28, 2021).
  - [83] A. Ghazi, P. Saba, *et al.*, "Utilizing 3D Printing and Hydrogel Casting for the Development of Patient-Specific Rehearsal Platforms for Robotic Assisted Partial Nephrectomies," *Urology*, vol. 147, p. 317, Jan. 2021, doi: 10.1016/j.urology.2020.10.023.
  - [84] B. Gore, "NASA TLX: Task Load Index," *NASA*, Dec. 15, 2020.  
<https://humansystems.arc.nasa.gov/groups/TLX/> (accessed Feb. 17, 2021).
  - [85] S. Said, M. Gozdzik, *et al.*, "Validation of the raw national aeronautics and space administration task load index (NASA-TLX) questionnaire to assess perceived workload in patient monitoring tasks: Pooled analysis study using mixed models," *J. Med. Internet Res.*, vol. 22, no. 9, Sep. 2020, doi: 10.2196/19472.
  - [86] R. A. Grier, "How high is high? A meta-analysis of NASA-TLX global workload scores," in *Proceedings of the Human Factors and Ergonomics Society*, Sep. 2015, vol. 2015–January, no. 1, pp. 1727–1731, doi: 10.1177/1541931215591373.
  - [87] C. J. Hodson, J. A. Drewf, *et al.*, "RENAL SIZE IN NORMAL CHILDREN A RADIOGRAPHIC STUDY DURING LIFE," doi: 10.1136/adc.37.196.616.
  - [88] S. G. Hart, "NASA-task load index (NASA-TLX); 20 years later," in *Proceedings of the Human Factors and Ergonomics Society*, Oct. 2006, vol. 50, no. 9, pp. 904–908, doi: 10.1177/154193120605000909.
  - [89] M. R. Wilson, J. M. Poolton, *et al.*, "Development and validation of a surgical workload measure: The surgery task load index (SURG-TLX)," *World J. Surg.*, vol. 35, no. 9, pp. 1961–1969, Sep. 2011, doi: 10.1007/s00268-011-1141-4.
  - [90] C. Kunz, P. Maurer, *et al.*, "Infrared marker tracking with the HoloLens for neurosurgical interventions," *Curr. Dir. Biomed. Eng.*, vol. 6, no. 1, May 2020, doi: 10.1515/cdbme-2020-0027.
  - [91] W. R. Crum, T. Hartkens, *et al.*, "Non-rigid image registration: Theory and practice," *British Journal of Radiology*, vol. 77, no. SPEC. ISS. 2. 2004, doi: 10.1259/bjr/25329214.
  - [92] A. Morelli, S. Moccia, *et al.*, "Towards deformable registration for AR in nephrectomy," 2020.

Accessed: Feb. 04, 2021. [Online]. Available:  
<http://research.intusurg.com/dvrkwiki/index.php?title=Main>.

- [93] N. Golse, A. Petit, *et al.*, “Augmented Reality during Open Liver Surgery Using a Markerless Non-rigid Registration System,” *J. Gastrointest. Surg.*, Feb. 2020, doi: 10.1007/s11605-020-04519-4.
- [94] Y. Adagolodjo, N. Golse, *et al.*, “Marker-Based Registration for Large Deformations - Application to Open Liver Surgery,” in *Proceedings - IEEE International Conference on Robotics and Automation*, Sep. 2018, pp. 4007–4012, doi: 10.1109/ICRA.2018.8462909.



## 12 Appendix

### 12.1 UMBRELLA staging criteria

#### Stage I

- Tumour is limited to the kidney.
- Tumour is present in the perirenal fat but is surrounded by a fibrous (pseudo)capsule. The (pseudo)capsule might be infiltrated by viable tumour, which does not reach the outer surface.
- Tumour might show protruding (botryoid) growth into the renal pelvis or the ureter but does not infiltrate their walls.
- The vessels or the soft tissues of the renal sinus are not involved by tumour. Intrarenal vessel involvement might be present.

#### Stage II

- Viable tumour is present in the perirenal fat and is not covered by a (pseudo)capsule, but is completely resected (resection margins are clear).
- Viable tumour infiltrates the soft tissues of the renal sinus.
- Viable tumour infiltrates blood and/or lymphatic vessels of the renal sinus or of the perirenal tissue, but it is completely resected.
- Viable tumour infiltrates the wall of the renal pelvis or of the ureter.
- Viable tumour infiltrates the vena cava or adjacent organs (except the adrenal gland) but is completely resected.

#### Stage III

- Viable tumour is present at a resection margin. Nonviable tumour or chemotherapy-induced changes present at a resection margin are not regarded as stage III.
- Abdominal lymph node involvement is present by either viable or nonviable tumour.
- Preoperative or intraoperative tumour rupture, if confirmed by microscopic examination (viable tumour at the surface of the specimen at the area of the rupture).
- Viable or nonviable tumour thrombus is present at resection margins of ureter, renal vein, or vena cava inferior (always discuss resection margins with the surgeon).
- Viable or nonviable tumour thrombus, which is attached to the inferior vena cava wall, is removed piecemeal by a surgeon.
- Wedge or open tumour biopsy before preoperative chemotherapy or surgery.
- Tumour implants (viable or nonviable) are found anywhere in the abdomen.
- Tumour (viable or nonviable) has penetrated through the peritoneal surface.

**Stage IV**

- Haematogenous metastases (for example, lung, liver, bone and brain) or lymph node metastases outside the abdominopelvic region.

**Stage V**

- Bilateral renal tumours at diagnosis. Each side should be substaged according to the above criteria.

## 12.2 Questionnaire clinical validation

**HoloLens kidney registration application questionnaire**

This part of the questionnaire is based on the NASA Task Load Index. It consists of six categories which are scaled in a 21 gradation scales. This part is designed to rates the perceived workload when using the HoloLens 2 application for NSS.

Name:	Function:	Date:
-------	-----------	-------

**Mental Demand** How mentally demanding was the task?

<div style="display: flex; justify-content: space-between; border-top: 1px solid black; border-bottom: 1px solid black; height: 20px;"> <span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span> </div>	<div style="display: flex; justify-content: space-between; border-top: 1px solid black; border-bottom: 1px solid black; height: 20px;"> <span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span> </div>
Very Low	Very High

**Physical Demand** How physically demanding was the task?

<div style="display: flex; justify-content: space-between; border-top: 1px solid black; border-bottom: 1px solid black; height: 20px;"> <span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span> </div>	<div style="display: flex; justify-content: space-between; border-top: 1px solid black; border-bottom: 1px solid black; height: 20px;"> <span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span> </div>
Very Low	Very High

**Temporal Demand** How hurried or rushed was the pace of the task?

<div style="display: flex; justify-content: space-between; border-top: 1px solid black; border-bottom: 1px solid black; height: 20px;"> <span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span> </div>	<div style="display: flex; justify-content: space-between; border-top: 1px solid black; border-bottom: 1px solid black; height: 20px;"> <span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span> </div>
Very Low	Very High

**Performance** How successful were you in accomplishing what you were asked to do?

<div style="display: flex; justify-content: space-between; border-top: 1px solid black; border-bottom: 1px solid black; height: 20px;"> <span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span> </div>	<div style="display: flex; justify-content: space-between; border-top: 1px solid black; border-bottom: 1px solid black; height: 20px;"> <span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span> </div>
Perfect	Failure

**Effort** How hard did you have to work to accomplish your level of performance?

<div style="display: flex; justify-content: space-between; border-top: 1px solid black; border-bottom: 1px solid black; height: 20px;"> <span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span> </div>	<div style="display: flex; justify-content: space-between; border-top: 1px solid black; border-bottom: 1px solid black; height: 20px;"> <span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span> </div>
Very Low	Very High

**Frustration** How insecure, discouraged, irritated, stressed, and annoyed were you?

<div style="display: flex; justify-content: space-between; border-top: 1px solid black; border-bottom: 1px solid black; height: 20px;"> <span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span> </div>	<div style="display: flex; justify-content: space-between; border-top: 1px solid black; border-bottom: 1px solid black; height: 20px;"> <span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span><span></span> </div>
---	---

Very Low

Very High

**The second part consists of multiple application questions in Likert scale (1 strongly disagree -5 strongly agree). This part is designed to get quantitative answers which can be easily compared to the answers of the other surgeons. The questions in this part will give an insight of your experiences of the application.**

The user-interface is easy and intuitive

1                      2                      3                      4                      5

The QR-code recognition worked immediately

1                      2                      3                      4                      5

The QR-tracking was performed sufficient during the procedure

1                      2                      3                      4                      5

The anatomic landmarks were easy and were interpreted intuitively

1                      2                      3                      4                      5

The anatomic landmarks were easily set onto the phantom

1                      2                      3                      4                      5

The virtual content was accurately overlaid onto the phantom kidney

1                      2                      3                      4                      5

The virtual overlay did not hindered the view of the surgical site

1                      2                      3                      4                      5

I was satisfied with how the abdominal phantom and testing environment resembled a real surgical setting

1                      2                      3                      4                      5

I was satisfied with how the content was visualised

1                      2                      3                      4                      5

I am satisfied with the time it took to successfully run through the application

1                      2                      3                      4                      5

This application has an added value for localizing the tumour

1                      2                      3                      4                      5

This application could be implemented into the clinical workflow of a NSS

1                      2                      3                      4                      5

I would use this application during a partial nephrectomy in the future

1                      2                      3                      4                      5

**Part three consists of some open questions to get an insight of your personal opinion of this application and your experiences with the application.**

How many times did you performed the application for each kidney phantom to get the best possible overlay?

Phantom 1      \_\_\_\_\_

Phantom 2      \_\_\_\_\_

Phantom 3      \_\_\_\_\_

Phantom 3      \_\_\_\_\_

Did you recognize a personal learning curve during this experiment?

---

Was there a difference in accuracy in a left or right kidney, if so what was the difference?

---

Was there a difference in accuracy with a large or small tumour, if so what was the difference?

---

Do you think this application is suitable for clinical implementation during NSS? Explain your answer

---

Do you see other surgical procedures as a future possibility for using the HoloLens? If so, which procedures?

---

Do you have any improvement suggestions?

---

Additional comments:

### 12.3 Abstract CARS 2021

# Towards holographic navigation for paediatric nephron sparing Wilms tumour surgery: a technical feasibility study

**Purpose:** A small part of patients suffering from Wilms tumours (WT) can be treated with open nephron sparing surgery (NSS). During this procedure intraoperative ultrasound is used to visualise the exact size of the tumour. Unfortunately, approximately 30% of the NSS procedures still result in a positive surgical margin. New visualisation techniques may help us reduce this percentage.

The size and infiltration of the tumour can be visualised intraoperatively with new techniques such as holographic navigation based on Augmented Reality (AR). The HoloLens 2 projects an overlay of the preoperative imaging in AR onto the open surgical field. Using AR surgical navigation technology, our goal is to reduce positive surgical margins during NSS. Therefore, we have developed an AR application which determines the required overlay with a combination of two algorithms; a registration algorithm based on anatomical landmarks and a tracking algorithm based on QR-code recognition. These self-developed algorithms are validated in this feasibility study. The clinically maximally allowed error for this validation was determined to be five mm by our surgical team.

**Methods:** The validation of this self-developed AR application is twofold. We measure the accuracy of the tracking through QR-code recognition and we estimate the accuracy of the resulting registration based on a 3D-printed kidney phantom.

In order to measure the accuracy of the QR-code recognition, a calibration setup was built in which the HoloLens was placed onto a 3D-printed bearer and the QR-code could be moved manually in two different directions. The HoloLens 2 measured the moved distance of the QR-code, which was compared with the actual movement. Additionally, we investigated the minimal workable size of the QR-code without limiting the tracking quality. Three different sizes were used, 5x5 cm, 4.5x4.5 cm, and 4x4 cm.

The accuracy of the registration algorithm was derived through a phantom study in which seven volunteers participated. None of our volunteers had prior experience with the AR application. A short introduction on how to use the HoloLens 2 was given and eye-calibration was performed. The volunteers were asked to perform the registration by pinpointing five anatomical landmarks onto the kidney phantom using a surgical pointer, as is shown in figure 1. After mapping of the anatomical landmarks, the holographic 3D model is automatically registered onto the 3D printed phantom. Subsequently, six randomly placed positional markers are highlighted on top of the holographic visualisation. The volunteers were asked to mark these holographic positional markers in the real world with a marker. These marked positions were compared with the actual position using a 3D-printed calibration mold, which allowed us to quantify the amount of misalignment between the holographic 3D model and the corresponding 3D printed model.

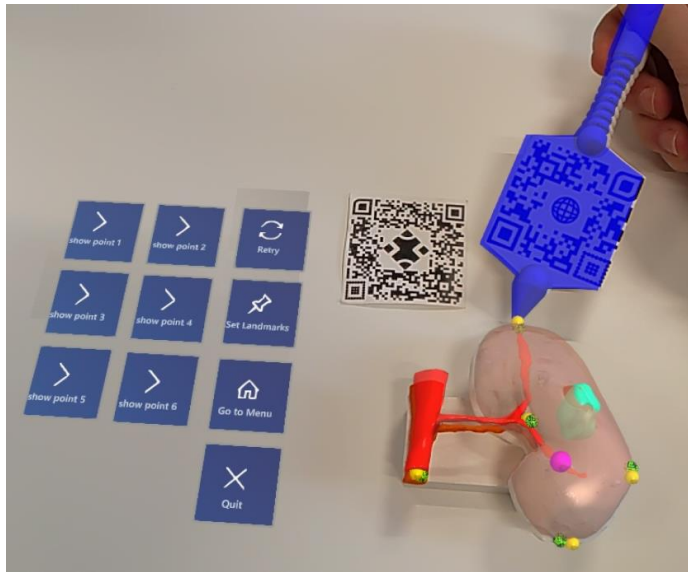


Figure 6, The holographic overlay of the kidney (salmon), tumour (cyan), artery (red), and a positional marker (purple) on top of the 3D printed model are shown. Additionally, the QR-code used for tracking and the QR-code placed onto the surgical pointer (blue) are visible. The surgical pointer is used to pinpoint the anatomical landmark for the HoloLens 2. There is a slight offset of the overlay due to a difference in height of the camera of the HoloLens 2 and the eyes of the user.

**Results:** We measured the movement of a moving QR-code with three different dimensions (5x5, 4.5x4.5 and 4x4 cm) in two directions. The mean error and the standard deviation of the movement in the X-axis was 0.29 mm (2.2mm), 4.04 mm (4.6), and 11.54 mm (3.8) respectively. For the movement in the Z-axis the mean error and standard deviation was 0.47 mm (3.6), 4.18 mm (5.6), and 13.57 (8.2) respectively. There was a significant difference in mean error between the 5x5 cm and 4.5x4.5 cm QR-codes and the 5x5 cm and 4x4 cm QR-codes ( $p < 0.001$ ).

The accuracy of the registration was measured as a mean distance between the holographic and actual position. This distance varied between volunteers but these differences were statistical not significant ( $0.308 \leq p \leq 0.963$ ). Unfortunately, one positional marker was excluded from the measurements due to a complete misalignment of the holographic marked position and actual position. The mean distance error per volunteer is given in table 1. Three of the 35 holographic marked positions were placed with a distance greater than 5 mm of the actual position by two different volunteers. Only one volunteer obtained a mean error above the clinical error of five mm.

Table 7 the mean distance and standard deviation of all 5 marked positions per volunteer

Volunteer	Mean error [mm] (Std. Deviation)
1	2.80 (0.99)
2	2.75 (0.92)
3	2.87 (1.11)
4	3.01 (0.95)
5	2.27 (1.60)
6	5.19 (4.30)
7	3.93 (2.19)

**Conclusion:** The accuracy obtained by our surgical navigation AR application is within the clinically allowed error and therewith we believe this technique may be feasible for clinical use. The results show that the QR-code tracking capacity of our application is sufficient and that volunteers are capable to apply the anatomical landmark based registration satisfactory. Further implementation of



this technique during nephron sparing surgery is recommended. However, intraoperative surgical workload and useability needs to be assessed beforehand. Therefore, we recommend performing a surgical phantom study to further explore this technique in paediatric oncologic surgery.