

# Texturextraktion aus endoskopischen Videoaufnahmen für die virtuelle Endoskopie am Beispiel der Trachea

---

M. Neugebauer<sup>1</sup>, R. Gasteiger<sup>1</sup>, C. Arens<sup>2</sup>, L. Dornheim<sup>1</sup>, B. Preim<sup>1</sup>

<sup>1</sup> Otto-von-Guericke-Universität, Institut für Simulation und Graphik, Magdeburg, Deutschland

<sup>2</sup> Universitätsklinikum, Klinik für Hals-, Nasen- und Ohrenheilkunde, Magdeburg, Deutschland

Kontakt: mathias.neugebauer@ovgu.de

## Abstract:

Die virtuelle Endoskopie besitzt ein breites Anwendungsspektrum, sei es für die Falldokumentation, Lehrsysteme oder die Planung von Eingriffen. Den Schichtbildern (CT, MR) als Grundlage, liegen aber nur geometrische Informationen ohne qualitative Informationen über die Beschaffenheit der Organwand vor. Am Beispiel der Trachea wurde im Rahmen dieser Arbeit untersucht, wie Texturen aus endoskopischen Videoaufnahmen extrahiert und für die virtuelle Endoskopie genutzt werden können. Als Datengrundlage dienen CT-Schichtbilder und Endoskopie-Videos aus denen das virtuelle Trachea-Modell und die Texturen gewonnen werden. Für die Texturextraktion nutzen wir projektive Texturierung und Textur-Stitching. Neben patientenspezifischen Daten dient ein experimenteller Aufbau zur genaueren Untersuchung des Einflusses von Aufnahme-Parametern. In den vorläufigen Ergebnissen stellt sich die bildbasierte Texturextraktion als vielversprechend dar, der Einsatz spezialisierter Aufnahme-Hardware ist aber wahrscheinlich notwendig, um das Korrespondenzproblem zwischen Videoaufnahme und 3D-Modell zu lösen.

Schlüsselworte: virtuelle Endoskopie, Trachea, Texturierung

## 1 Problemstellung

Bei der virtuellen Endoskopie (VE) wird aus Schichtbildern (z.B. CT oder MR) ein dreidimensionales Modell des betreffenden Organs (Darm, Trachea, etc.) rekonstruiert. Dann kann an einem Computer mithilfe einer virtuellen Kamera das Organ von innen exploriert werden [1]. Es kann also eine Endoskopie durchgeführt werden, ohne dass der Patient diesem invasiven Eingriff ausgesetzt werden muss. Weiterhin kann das Organ aus Blickwinkeln exploriert werden, die aufgrund räumlicher oder technischer Begrenzungen in einer realen Untersuchung nicht möglich sind. Durch entsprechende Visualisierungssysteme können Oberflächenbesonderheiten hervorgehoben und zusätzliche Informationen, wie z.B. visuell verdeckte, naheliegende Risikostrukturen, eingeblendet werden. Virtuelle Werkzeuge ermöglichen eine einfache Vermessung von Abständen und Größen und wenn die VE mit einem entsprechenden Simulationssystem gekoppelt ist, können sogar ganze Eingriffe am Computer erprobt werden.

Aufgrund dieser vielfältigen Möglichkeiten besitzt die VE ein breites Anwendungsspektrum [2]. Sie kann für die Patientenaufklärungen oder für die multimedial erweiterte Falldokumentation genutzt werden. Des Weiteren kann sie im Rahmen von Lehrsystemen für das Training endoskopischer Eingriffe zum Einsatz kommen und zusätzliche Informationen für die Therapieplanung und -durchführung liefern [3] [4].

Allerdings hat die klassische VE einen entscheidenden Nachteil. Da als einzige Datenquelle Schichtbilder herangezogen werden, stehen qualitative Informationen über die Beschaffenheit des Organgewebes nur im geringen Maße, meist aber gar nicht zur Verfügung. Am Beispiel der Trachea können z.B. die einzelnen Spangen sehr gut räumlich aus den Schichtbildern rekonstruiert werden, eine Aussage über entzündete Gewebereiche kann allerdings nicht getroffen werden. Deswegen kann eine VE eine echte Endoskopie weniger ersetzen, sondern durch die oben beschriebenen, zusätzlichen Möglichkeiten ergänzen.

Ziel dieser Arbeit ist es zu untersuchen, wie aus Filmaufnahmen einer Trachea-Endoskopie Texturen für die Gewebeabbildung in der VE extrahiert werden können. Die vorgestellten Konzepte zielen dabei auf die Anwendung einer ergänzenden Falldokumentation (z.B. für eine spätere Diskussion mit Kollegen) und Patientenaufklärung ab. Notwendig dafür ist eine Endoskopie vor dem eigentlichen endoskopischen Eingriff, wor die reale Szene möglichst vollständig aufgenommen wird. Wir konzentrieren uns auf drei übliche Endoskopiewinkel: 0° (in Richtung der *Bifurcatio Trachea*), 60° (schräg zur Trachea-Wand) und 90° (direkt auf die Trachea-Wand). Ausgehend von diesen Aufnahmen ergeben sich zwei mögliche Texturierungsverfahren: projektive Texturierung und *Stitching*-Texturen. Parallel zur Untersuchung auf

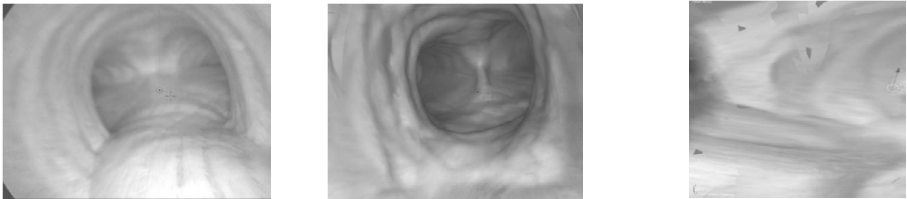
realen Patientendaten wurde ein experimenteller Aufbau geschaffen, der eine genauere Untersuchung des Einflusses verschiedener Aufnahmeparameter auf die Texturqualität erlaubt.

## 2 Material und Methoden

Als Datengrundlage stehen uns CT-Aufnahmen und endoskopische Videos ( $0^\circ, 60^\circ, 90^\circ$ ) einer Trachea zur Verfügung. Aus den CT-Schichtbildern muss ein 3D Modell der Trachea rekonstruiert werden. Aus den Filmaufnahmen müssen *Frames* extrahiert und aufbereitet werden, damit sie für die Texturerzeugung geeignet sind. Dann wird mithilfe der beiden Texturierungsmethoden ein entsprechend texturiertes 3D Modell für die VE erzeugt. Die Ergebnisse wurden anhand eines Direktvergleichs mit den endoskopischen Filmaufnahmen verglichen. Parallel dazu wurde derselbe Prozess auf einer künstlichen Trachea durchgeführt.

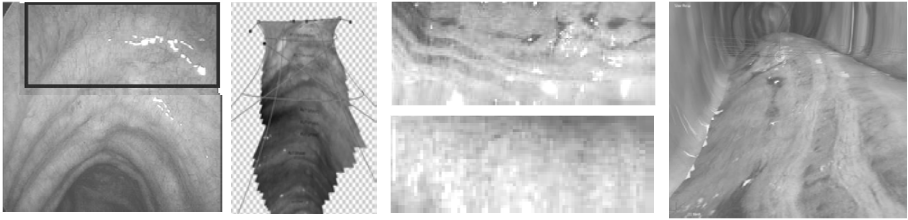
**Rekonstruktion:** Für die Segmentierung der Trachea wird ein *Region Growing* angewendet. Da sie luftgefüllt ist, weist sie in den CT-Aufnahmen eine homogene Intensität und eine gute Abgrenzung zum umgebenden Gewebe auf. Eine maximal erlaubte Intensitätsabweichung von 1,3% hat sich als gute Parametrisierung erwiesen. Die resultierende Segmentierungsmaske wird erweitert (Dilatation mit  $3 \times 3$  Kernel) und auf den invertierten Original-Datensatz multipliziert. Nun sind nur noch die Trachea und ein kleiner Sicherheitsrand in den Bilddaten vorhanden. Die Bildintensitäten werden zwischen 0 und 1 skaliert. Nun kann der *Marching Cubes* Algorithmus angewendet werden, um aus dem maskierten Bild ein Oberflächenmodell der Trachea zu erzeugen. Ein Iso-Wert von 0.94 liefert ein gutes Ergebnis, mit einer detaillierten Abbildung der Trachea-Spangen (**Abb. 1, Mitte**). Unerwünscht segmentierte Bereiche (z.B. Mundraum, Bronchialäste) werden in einem 3D-Editor (Blender) mit Standardwerkzeugen entfernt.

**Video-Aufbereitung:** Die zur Verfügung stehenden Videoaufnahmen (*Olympus Exera II System*, Videoverarbeitung: *rpScene HD*, komprimiert auf  $720 \times 544$ , MP4) bestehen aus Halbbildern, d.h. ein Videoframe besteht aus zwei Bildern, die gerade *Scan-Lines* aus dem ersten und die ungeraden aus dem zweiten Bild. Dies führt zu horizontalen Streifenartefakten. Um einen *Video-Frame* als Textur nutzbar zu machen, muss ein *De-interlacing* durchgeführt werden, bei dem die ungeraden *Scan-Lines* über Interpolation der Farbwerte aufgefüllt werden (**Abb. 1, links**). Dies wurde mit *VirtualDub* umgesetzt. Des Weiteren führt die Optik des Endoskops zu einer Kissenverzerrung. Diese wurde über ein Testbild mithilfe eines *Bezier-Gitters* modelliert. Ein daraus erzeugtes Entzerrungsgitter wird auf alle Frames angewendet.



**Abbildung 1:** Endoskopie-Frame (*links*), texturierte, virtuelle Trachea (*Mitte*), Verzerrungsartefakte und Unschärfe bei projektiver Texturierung (*rechts*)

**Projektive Texturierung:** Für die projektive Texturierung muss die Ausrichtung und Position der virtuellen Kamera in der virtuellen Trachea der echten Kamera aus den endoskopischen Aufnahmen entsprechen. Für diese korrelierende Ausrichtung wird eine Landmarke benötigt. Dazu können die *Bifurcatio Trachea*, die Aufspaltung am Ende Trachea, sowie auffällige Spangen genutzt werden. Im aktuellen Verfahren erfolgt die Wahl der Landmarken und die Ausrichtung der virtuellen Kamera manuell. Zwischen den ausgerichteten Kamerapositionen wird linear interpoliert und ggf. weitere Positionen eingefügt, falls die Abweichung zum realen Kamerapfad zu stark ist. Stimmen virtuelles und reales Endoskopiebild überein, wird eine lokale, projektive Parametrisierung erzeugt. Dann wird der aktuelle *Video-Frame* als Textur verwendet. Die Trachea ist so allerdings nur aus dem aktuellen Blickwinkel visuell korrekt texturiert. Ändert sich der virtuelle Blickwinkel, sieht man deutlich Verzerrungen und falsch texturierte Bereiche (**Abb. 1, rechts**). Um eine vollständige Texturabdeckung zu erreichen, wird aus weiteren Blickwinkeln lokal projektiv parametrisiert und texturiert. Die lokalen Texturen werden über manuelles *Textur-Painting* in eine globale Atlas-Textur mit einer globalen Oberflächenparametrisierung (Blender: Smart-UV-Project, Feature-Winkel:  $66^\circ$ ) überführt.



**Abbildung. 2:** Endoskopie-Frame mit relevantem Ausschnitt (links), Stitching-Textur während Entzerrung (Mitte links), Ausschnitte aus finalelem Texturstreifen (Mitte rechts), texturierte Trachea (rechts)

**Textur-Stitching:** Ein weiterer Ansatz ist es, Teile aus den Frames der Endoskopie-Aufnahme auszuschneiden, zu einer Gesamttextur zusammenzufassen und diese dann auf das 3D Modell der Trachea zu übertragen. Pro Frame ist der Teil von Interesse, in dem die Wand detailliert abgebildet ist. Bei  $90^\circ$  Aufnahmen ist es der gesamte Frame. Bei  $60^\circ$ -Aufnahmen, kann aufgrund der perspektivischen Verkürzung entlang der Trachea-Röhre, nur ein peripherer Teil des Frames genutzt werden. Bei  $0^\circ$ -Aufnahmen ist dieser Teil zu klein. Bei der Videoaufnahme wurde das Endoskop langsam und gradlinig in die Trachea eingeführt. Bei  $60^\circ$  und  $90^\circ$  Aufnahmen wurde dieser Prozess viermal wiederholt, wobei durch  $90^\circ$ -Rotation um die Endoskop-Achse jedes Mal ein anderer Streifen der Tracheawand aufgenommen wurde. Der jeweils interessante Framebereich wird visuell identifiziert und für alle Frames ausgeschnitten (**Abb. 2**, links). Über ein automatisches RANSAC Image Stitching werden die Ausschnitte zu einem Texturstreifen zusammengefasst. Die perspektivische Verkürzung in den  $60^\circ$ -Aufnahmen sorgt für eine Stitching-Textur bei der jeder neu hinzugefügt Teilabschnitt größer als der vorherige ist. Deswegen muss der Texturstreifen entsprechend entzerrt werden (ebenfalls durch ein Bezier-Gitter, **Abb. 2**, Mitte links). Um die Verzerrung zu modellieren, werden Landmarken in die Original-Frameabschnitte manuell eingefügt. Der bearbeitete Texturstreifen kann dann über multiple, projektive Texturierung auf die jeweilige Wandseite der virtuellen Trachea übertragen werden. Dafür wird eine orthographische Projektion verwendet, deren Blickwinkel zur Wand dem Aufnahmewinkel der Endoskopie entspricht.

**Experiment:** Um genauere Aussagen darüber treffen zu können, wie bestimmte Aufnahmeparameter die Texturqualität beeinflussen, wurde eine künstliche Trachea als zylindrische Röhre modelliert und mit einer Text-Textur ausgekleidet. Die Text-Textur besteht aus drei Bereichen: einem alphanumerisch annotiertem Koordinatengittern (Bereich A), einem Testfotobereich (Bereich B) und einer künstlich erzeugten Gewebetextur (Bereich C). Bereich A dient der Untersuchung der Textur-Verzerrung, Bereich B der Abbildung von Details und Bereich C der Klärung von Problemen, die durch selbstähnliche, detailarme Gewebeschnitte entstehen können. Mit einem Endoskop-Nachbau, der die Fixierung bestimmter Freiheitsgrade (Rotation, Bewegung) ermöglicht, wurden  $0^\circ$ ,  $60^\circ$  und  $90^\circ$ -Aufnahmen angefertigt. Dabei wurde die Beleuchtung, Schärfe, Bewegungs- und Rotationsgeschwindigkeit der Kamera variiert.

### 3 Ergebnisse

Für die Rekonstruktion standen zwei Trachea-Datensätze zur Verfügung. Beide konnten mit der oben beschriebenen Methode problemlos segmentiert werden. Bei dem ersten Datensatz stellte sich die Trachea mit einem runden Querschnitt dar. Dieser entspricht nicht dem Trachea-Querschnitt aus den Videoaufnahmen. Dort wurden immer eine abgeflachte, muskuläre Seite (zur Speiseröhre hin) und ein runder, durch die Spangen stabilisierter Bereich beobachtet. Der zweite Trachea-Datensatz wies diesen charakteristischen Querschnitt auf, allerdings gab es im mittleren Bereich einen Versatz der Schichtbilder. Dies ist auf eine Bewegung des Patienten während der Aufnahme zurückzuführen und führt zu einer deutlich sichtbaren Stufe im rekonstruierten Trachea-Modell.

Diese durch Artefakte herbeigeführte, schwache Korrespondenz zwischen virtueller und realer Trachea erschwerte das Herstellen korrespondierender Blickwinkel für die projektiven Texturierung. Nur Teile der Frames konnten über die Platzierung der Kamera in der virtuellen Trachea nachgeahmt werden. Dies war auch nur für die  $0^\circ$ -Aufnahmen möglich. Bei den  $60^\circ$ - und  $90^\circ$ -Aufnahmen war zu wenig von der Trachea zu sehen, um die virtuelle Kamera verlässlich auszurichten. Trotzdem konnte aus den  $0^\circ$ -Aufnahmen für einen Teil der Trachea eine flächendeckende Gesamttextur erzeugt werden. Diese war allerdings teilweise stark verzerrt und wies eine deutliche Unschärfe auf. Die Bildqualität der Stitching-Textur war deutlich besser. Allerdings konnte diese nur für  $60^\circ$ - und  $90^\circ$ -Aufnahmen erzeugt werden (**Abb. 2**, Mitte rechts). Bei  $0^\circ$ -Aufnahmen war der periphere Bildbereich zu klein, um überlagernden Texturteile für das Stitching bereitzustellen. Bei  $90^\circ$ -Aufnahmen wieder zeigten sich verstärkt Glanzartefakte, da das Endoskop-Licht direkt auf die feuchte Trachea-Wand schien. Außerdem waren einige Frames sehr unscharf, da die Kamera während der  $90^\circ$ -Aufnahmen zu dicht an die Trachea-Wand geführt wurde. Auch wenn ein größerer Bereich der Wand in der Stitching-Textur sichtbar war, war es nicht immer möglich, ein korrekte Platzierung der Textur auf der virtuellen Tra-

chea sicherzustellen. Während die Spangen deutlich im 3D-Modell sowie der Textur hervortraten und eine Ausrichtung ermöglichten, war der muskuläre Bereich sehr glatt, so dass dieser Texturstreifen nicht eindeutig platziert werden konnte (Abb. 2, rechts).

Bei den Experimenten konnten gute Ergebnisse erreicht werden, auch wenn die Kamera schnell bewegt oder nicht fokussiert war. Auch die Beleuchtung hatte keinen starken Einfluss auf die Ergebnistextur. Den stärksten Einfluss hatte der Aufnahmewinkel, mit ähnlichen Effekten wie bei den realen Endoskopieaufnahmen. Außerdem stellte die Selbstähnlichkeit des alphanumerischen Koordinatengitters und der Gewebetextur ein Problem dar. Beim Stitching wurden nicht benachbarte Teiltexturen falsch zusammengefügt. Dieser Effekt konnte allerdings nicht bei den realen Aufnahmen beobachtet werden. Diese hatten zwar homogen wirkende Bereiche, kleine Gefäße in der Trachea-Wand erlaubten allerdings trotzdem eine korrekte Texturzuordnung.

## 4 Diskussion

Ausgehend von den vorläufigen Ergebnissen zeigt sich, dass ein vollständig bildbasierter Ansatz auf Basis einer einzelnen Kamera für die Texturerzeugung nicht zu empfehlen ist. Das größte Problem ist die Korrespondenz zwischen Videoaufnahme und virtueller Trachea. Die Position der Endoskopkamera muss dazu implizit aus den Videoaufnahmen abgeleitet werden. Dies gelingt am besten bei den 0°-Aufnahmen, da diese in jedem Frame einen großen Teil der Trachea zeigen. Zusätzlich ist die Bifurkation als wichtige Landmarke in vielen Frames sichtbar. Allerdings sind 0°-Aufnahmen nur für projektives Texturieren geeignet. Da aufgrund der perspektivischen Verkürzung ein kleiner Bildbereich aus den Video auf einen vergleichsweise große Wandbereich des virtuellen Modells übertragen wird, ist die Ergebnistextur unscharf und starke Verzerrungen treten deutlich hervor. Beim *Stitching* ist die Bildqualität besser, weil die Gesamtextur aus detaillierten Teilbereichen der einzelnen Aufnahmen aufgebaut wird. Allerdings ist es dort trotz einer größeren Textur aufgrund der sich wiederholenden (Spangen) oder nicht vorhandenen (Muskelbereich) Landmarken schwierig, festzustellen, wie die Textur zu platzieren ist. Eine Lösung für das generelle Korrespondenzproblem ist der Einsatz eines Trackingsystems wie z.B. eine *Polaris*. Entsprechend kalibriert kann sie Informationen über Position und Lage der Kamera zu jedem Video-Frame bereitstellen. Sollte ein solches *Tracking*-System aufgrund der räumlichen Gegebenheiten im OP-Saal nicht eingesetzt werden können, wäre auch der Einsatz von zwei Endoskopie-Kameras denkbar. Eine 0°-Aufnahme dient zur räumlichen Orientierung und Korrespondenzerstellung, während eine zeitgleich durchgeführte 90°-Aufnahme die detaillierten Texturen liefert. Dies würde auch den Aufnahmeprozess erleichtern und so unscharfen oder fehlenden 90°-Aufnahmen entgegenwirken. Des Weiteren wäre der Einsatz einer Führungsschiene denkbar, um die Abweichung von der Trachea-Mittelachse zu verhindern und den Aufnahmeprozess noch weiter zu erleichtern.

Eine weitere Facette des Korrespondenzproblems ist die geometrische Abweichung zwischen der rekonstruierten und der in den Videoaufnahmen abgebildeten Trachea. Der Patient ist in beiden Aufnahmen in verschiedenen Zuständen, was zu einer veränderten Trachea-Form führen kann. Zusätzlich sind Bewegungsartefakte, die in den Schichtaufnahmen kaum auffallen, deutlich im rekonstruierten Oberflächenmodell der Trachea zu sehen. Somit stammt die aus den Aufnahmen extrahierte Trachea-Textur von einer geometrisch abweichenden Trachea. Dies kann dazu führen, dass die Textur nicht eindeutig platziert werden kann oder das die geometrische Diskrepanz visuelle Artefakte erzeugt und Fehlinterpretationen begünstigt. In diesem Fall wäre es besser, auf patientenindividuelle Geometrie zu verzichten und stattdessen einen generischen Zylinder anzubieten. Die geometrische Diskrepanz zwischen realer Anatomie und VE-Geometrie wäre leichter nachzuvollziehen. Formvariationen der Trachea-Wand könnten dann nur visuell aus der Textur entnommen werden, was ein Übersehen von Auffälligkeiten begünstigen könnte.

Ist der Einsatz patientenindividueller Geometrie zwingend erforderlich, sollte diese nicht aus Schichtbildern, sondern direkt während der Endoskopie gewonnen werden. Diese wäre durch eine zusätzliche, Laser-basierte Tiefenkamera denkbar. Ähnlich wie die Textur könnten die gewonnenen, lokalen Tiefeninformationen zu einem globalen Trachea-Modell zusammengefasst werden. Eine vergleichbare Technik kommt beim *KinectFusion-Projekt* zum Einsatz, das als Hardware eine Tiefenkamera verwendet. Eine entsprechende Miniaturisierbarkeit der Kamertechniken ist potentiell möglich, wie in Maier-Hein et al. beschrieben [5].

**Fazit:** Auch wenn bildbasierte Ansätze mit impliziter Positionsbestimmung der Endoskopkamera vielversprechende Ergebnisse für die VE-Texturierung zeigen, ist die ungenaue Korrespondenz zwischen Videoaufnahme und Trachea-Modell problematisch. Ein Lösungsansatz könnte der Einsatz angepasster Hardware sein, beispielsweise in Form von *Tracking*-Systemen, multiplen Kameras oder der 3D-Rekonstruktion während der endoskopischen Aufnahme.

## 5 Referenzen

- [1] Arno Krüger et al., *Sinus endoscopy - application of advanced GPU volume rendering for virtual endoscopy*, IEEE Transactions on Visualization and Computer Graphics, 1491-1498, 2008

- [2] André Neubauer, Stefan Wolfsberger, *Virtual endoscopy in neurosurgery: a review*, *Neurosurgery* 72. Supplement 1 A97-A106, 2013
- [3] Arnulf Ferlitsch, et al., *Effect of virtual endoscopy simulator training on performance of upper gastrointestinal endoscopy in patients: a randomized controlled trial*, *Endoscopy* 42.12: 1049, 2010
- [4] Junchen Wang, et al., *Intravascular catheter navigation using path planning and virtual visual feedback for oral cancer treatment*, *The International Journal of Medical Robotics and Computer Assisted Surgery* 7.2: 214-224, 2011
- [5] Maier-Hein et al., *Optical techniques for 3D surface reconstruction in computer-assisted laparoscopic surgery*, *Medical Image Analysis*, 17(8), 974-996, 2013