Vergleich von quantitativen 3D-Messungen in der klinischen Endoskopie am Beispiel des Kehlkopfes

Thomas Wittenberg♠, Stefan Schuberth♣, Klaus Spinnler♠,
Robert Schmidt♠ und Ulrich Eysholdt♣

♠Fraunhofer Institut für Intergrierte Schaltungen, Erlangen
♣Abteilung für Phoniatrie und Pädaudiologie am Klinikum der Univ. Erlangen
Email: {wbg,spk,sch}@iis.fhg.de, stefan.schuberth@gmx.de

Zusammenfassung. In dieser Arbeit werden zwei unterschiedliche Ansätze zur Vermessung intrakorporaler Hohlräume mittels Endoskopen am Beispiel des Kehlkopfes gezeigt und miteinander verglichen. Der erste Ansatz aus der Gruppe der passiven Methoden ist ein Stereo-Ansatz und beruht auf einer kalibrierten Stereoansicht der abgebildeten Szene. Aus der Gruppe der aktiven Methoden, bei dem spezielle Beleuchtungen zum Einsatz kommen, wird ein Ansatz mit parallelen Laserstrahlen vorgestellt.

1 Motivation

Das Interesse an der — möglichst In-Situ — Quantifizierung von endoskopischen Untersuchungen zur Vermessungen morphologischer und funktioneller Veränderungen menschlicher Hohlräume zu diagnostischen und therapeutischen Zwecken hat in jüngster Zeit zugenommen. Aufgrund der meist sehr beengten Verhältnisse der zu untersuchenden Körperöffnungen (Bauchraum, Darm, Kehlkopf und Speiseröhre, ...) sind solche Techniken in der klinischen Endoskopie jedoch bisher kaum verbreitet.

Unter diesem Hintergrund wurden am Beispiel des menschlichen Kehlkopfes Untersuchungen durchgeführt, um dessen Komponenten (Stimmritze, Stimmlippen, Stellknorpel, ...) zu vermessen sowie die räumlichen Bewegungsabläufe der Stimmbänder quantifizieren zu können.

Ziel dieser Untersuchung war die Quantifizierung von Stimmlippenschwingungen auf der Basis von endoskopischen Aufnahmen mit einer Hochgeschwindigkeitskamera. Während sich die zeitliche Diskretisierung aufgrund der geeichten Kamera-Aufnahmegeschwindigkeit ermitteln lässt, können die räumlichen Aspekte dieser endoskopischen Aufnahmen bisher lediglich auf der Basis von relativen Pixelmaßen berechnet und dargestellt werden.

2 Material und Methode

Für die drei-dimensionale Erfassung und Vermessung von Objekten unter Verwendung von Kameras können zwei Ansätze unterschieden werden, namentlich

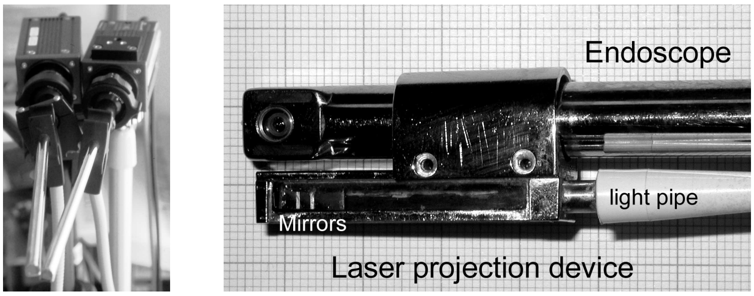

Abb. 1. *Stereo-Hochgeschwindigkeitskamera mit Doppelkamerakopf und zwei Endoskopen (links), Endoskop mit Laserprojektionsaufsatz (rechts)*

die sog. *passiven* und *aktiven* Verfahren, abhängig von der Art der Bilderfassung [1]. 3D-Messungen bzw. Aufnahmen werden als *aktiv* bezeichnet, wenn die zu untersuchende Szene mit zusätzlichem (Spezial-)Licht kontrolliert werden muss. Im Gegensatzsatz dazu benötigen *passive* Methoden neben den Kameras (im einfachsten Fall eine einzige Kamera mit Positionsmesser) keine weiteren Informationen oder spezielles Zubehör.

Für die vorliegende Untersuchung wurden zwei unterschiedliche Ansätze unter Verwendung von Endoskopen getestet, evaluiert und miteinander verglichen: Zum einen aus der Gruppe der passiven Ansätze ein Stereoverfahren, und zum anderen ein aktives Verfahren auf der Basis von Laserpunktprojektion.

2.1 Stereoansatz

Beim Ansatz mit dem Stereoverfahren wurde ein Stereo-Hochgeschwindigkeitskamerasystem mit zwei identischen Kameraköpfen (256×256 Pixel), identischen Optiken (17 mm) und identischen 90^o-Lupenendoskopen in einer festen Geometrie miteinander verbunden, vgl. Abb. 1. Die beiden Endoskope besaßen in dieser Anordnung einen Winkel von $\alpha = 20^o$ in der XY-Ebene und einen Winkel von $\beta = 150^o$ in der YZ-Ebene zueinander.

Um eine Relation zwischen den internen Kamerakoordinatensystemen und dem externen Weltkoordinatensystem herzustellen, müssen die verwendeten optischen Komponenten des Systems (Kameras, Sensoren, Optiken und Lupenendoskope) kalibriert werden. Zur Kalibrierung der intrinsischen und extrinsischen Parameter wurde der iterative Ansatz von Tsai [3]unter Verwendung eines koplanaren Kalibriermodells (Karomuster mit 5 mm Linienabstand) eingesetzt, da bei diesem Verfahren u.a. auch die radiale Verzerrung der verwendeten Linsen berücksichtigt wird. Da der Kameraaufbau bei dem Versuch konstant gehalten werde musste, wurde das Kalibriermuster nach jeder Aufnahme derart unter die beiden Lupenlaryngoskope platziert, bis die Messpunkte des Kalibriermodells in beiden Kameraansichten scharf abgebildet waren.

Nach dem Kalibrierprozess werden die beiden Kameraansichten miteinander verknüpft, indem semiautomatisch Punktkorrespondenzen an acht Stellen (Auf-

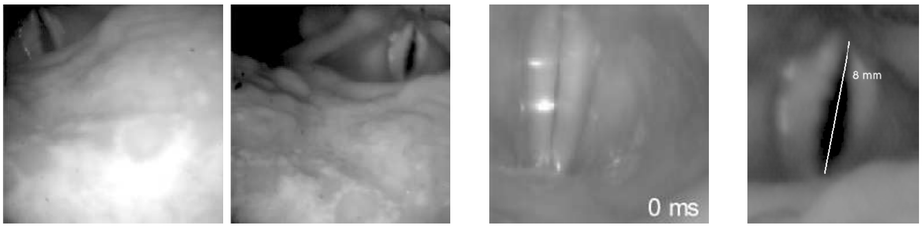

Abb. 2. *Stereo-Ansicht des Kelhlkopfs, linker und rechter Endoskopischer Anblick (links) Kehlkopf mit zwei projezierten Laserpunkten auf der rechten Stimmlippe (mitte), quantitative Erfassung der Glottislänge (rechts)*

punkten) auf den Stimmlippen in den Bildpaaren ermittelt werden, um einen 3D- bzw. Tiefendatensatz des Kehlkopfes zu zu erzeugen. Dieser Datensatz und die zugehörigen Trajektorien lassen sich nun in DIN-Einheiten (mm oder cm) messen (s. Abb. 3) [4].

2.2 Laserprojektion

Beim zweiten Ansatz wurde das 90^o-Lupen-Endoskop einer monokularen Hochgeschwindigkeitskamera um ein Laserprojektionssystem erweitert, mit dessen Hilfe zwei parallele Laserstrahlen im Abstand von 3.8 mm auf die Stimmlippen projiziert werden (s. Abb. 1) [2] . Durch die parallele Projektion der Strahlen wird der Abstand der projizierten Laserpunkte unabhängig von der Entfernung zwischen Projektionssystem und Projektionsebene, s. Abb. 2. Zusätzlich wurden die parallelen Laserstrahlen so eingestellt, dass diese nicht parallel zur optischen Achse des Aufnahmesystems, sondern unter einem Winkel von 6.8^o justiert wurden. Dieser Winkel leitet sich aus der klinischen Untersuchungsmethode ab, und es soll erreicht werden, dass die Laserstrahlen senkrecht auf die Stimmlippen treffen. Damit wird der Fehler minimiert, der durch das Kippen der Projektionsebene, in diesem Fall der Stimmlippen, entsteht. Durch den Winkel von 6.8^o zur optischen Achse ist es auch möglich, den Abstand der Stimmlippen zur Endoskopspitze zu bestimmen. Da die Laserstrahlen unter einem Winkel zur optischen Achse projiziert werden, bewegen sich diese im aufgenommenen Bild mit wachsendem Abstand zwischen Endoskopspitze und Projektionsebene nach außen. Somit kann jeder Position der Laserpunkte im Bild ein Abstand zwischen Endoskopspitze und Projektionsebene zugeordnet werden.

Durch eine Abstandsmessung der Laserpunkte im Bild ist es möglich, eine Skalierung zwischen relativen Pixel- und räumlichen DIN-Einheiten herzustellen. Die Laserpunkte werden dazu in jedem Bild automatisch detektiert, indem die hellsten Bereiche im Bild gesucht und eine zweidimensionale Gaussfunktion an diese Bildbereiche gefittet werden. So ist es möglich den Abstand der projizierten Laserpunkte subpixel genau zu extrahieren. Das Resultat ist ein Maßstab zwischen *Pixel* und *mm* in der Projektionsebene.

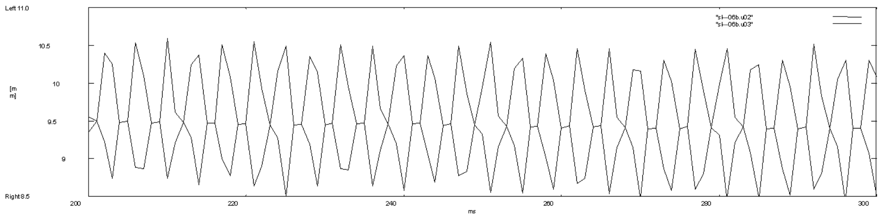

Abb. 3. *Kurzer Ausschnitt von* $100ms$ *der Bewegungstrajektorien der rechten und linken Stimmlippe in DIN-Massen (mm). Beide Seiten weisen Schwingungsamplituden in der Grössenordnung von* $1mm$ *auf*

Der Vorteil dieser Methode gegenüber dem Stereoansatzes liegt u.a. darin, dass neben einer einmaligen Kalibrierung des Laserprojektionssystems keine weitere Kalibriermessung nötig ist; die Kalibrierung ist zudem auch nach einem Wechsel der Optik gültig, da der Abstand der projizierten Laserpunkte konstant und bekannt ist.

3 Ergebnisse und Ausblick

Während sich mit beiden vorgestellten Verfahren die gewünschte räumliche Quantifizierung der endoskopischen Aufnahmen mit vergleichbaren Ergebnissen erreichen lässt (s. Abb. 2 und 3, Glottislänge $8mm$; Amplitude $1mm$), divergieren die Verfahren jedoch in Punkten der Handhabung, der klinischen Praktikabilität und der Flexibilität ihrer Erweiterungsmöglichkeiten. Obwohl der Stereoansatz die weitaus grössere Flexibilität besitzt und neben der Berechnung von quantifizierenden Grössen eine komplette 3D-Rekonstruktion der beobachteten endoskopischen Szene zulässt, ist dieser Ansatz derzeit aufgrund der zur Zeit fehlenden adäquaten optischen Komponenten sowie der komplexen Handhabung noch nicht ausgereift für einen routinemässigen Einsatz. Vielversprechender ist dagegen derzeit die Verwendung von mindestens zwei parallel-projizierten Laserstrahlen, die zwar keine vollständige 3D Rekonstruktion zulässt, aber sich aufgrund ihrer einfachen Handhabung sehr leicht in die klinische Routine einfügt.

Literatur

1. S. Posch. *Automatische Tiefenbestimmung aus Grauwertstereobildern.* Deutscher Univeristäts Verlag, Wiesbaden, 1990.
2. S. Schuberth, U. Hoppe, M. Döllinger, J. Lohscheller, T. Steudel, and U. Eysholdt. High-precision measurement of the vocal fold length and vibratory amplitudes. *The Laryngoscope*, angenommen.
3. R. Y. Tsai. An Efficient and Accurate Camera Calibration Technique for 3D Machine Vision. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 364–374, Miami Beach, FL, 1986.
4. T. Wittenberg, M. Tigges, K. Spinnler, and U. Eysholdt. Some thoughts about 3D and Stereo in Laryngoscopy. In *Proc. Advances in Quantitative Laryngoscopy, Voice and Speech Research*, pages 116–123, 2000.