# Visual Light related multi-modal sensor and computed structured data

## SUBMITTED BY \_\_\_\_\_\_\_\_\_\_\_\_Richard Bieck\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Presenter’s Name

## On Behalf of Working Group 24

Surgery

## Introduction/Scope

Over 80 % of today’s procedures are performed minimal-invasive using a form of endoscopic imaging. For certain types of clinical departments, e.g., in visceral surgery, numbers are even higher (GER, 90 %, US, 92 %). As a consequence, minimally-invasive technology is rapidly evolving to provide better imaging functionality, e.g., 3D vision through surgical microscopes and endoscopes to provide navigational aid in complex procedures. Each day, minimally-invasive procedures generate millions of gigabytes of surgical video data worldwide.

As Artificial Intelligence is advancing in medicine, this abundance of unprocessed large-scale data receives more and more attention. In comparison to any other known imaging format, the biggest challenge that such a large image base imposes is the extraction of relevant information. Three seconds of a video stream produce approximately the same amount of single interpretable images as a CT scan of the brain.

Not only does a single image contain semantic information that could have relevance for the overall treatment, e.g., the appearance of a specific anatomical landmark, but the video stream itself also represents a continuous sequence of image data to consider. The difference is that there are hardly more than 10.000 images in a current trans-sectional slide examination but millions in video data.

With image-based AI applications, the demand for sophisticated data content management and tools for the necessary labeling tasks will grow constantly. This matter becomes even more pressing when complex intraoperative navigation is based on streams of multi-sensor data and depth information derived from, e.g., VL stereoscopic imaging systems.

The underlying basic Use Case is straight-forward: During a surgical procedure 2D/3D video data as well as multi-modal sensor data is continuously acquired. This data is post-processed and can be used in several surgical scenarios.

Examples would be

* using the structured information for education by enhancing the video streams, e.g. with anatomical context information derived from image-based algorithms,
* using structured information from video streams for clinical decision support e.g. surgical workflow analysis, procedure documentation,
* using the structured information for scientific purposes, e.g. generation of large scale multi-modal datasets for training and benchmarking of machine learning applications.

Therefore, this work item proposal addresses the need to extend the DICOM Standard in a way that high quantities of acquired surgical navigation information, e.g., labels, multi-sensor data, annotations, and depth information, can be mapped to and stored along with the underlying 2D/3D video streams even at typical frequencies of 50+ Hz.

This proposal will specifically focus on the non-realtime use cases. We plan to focus on live use cases at a later stage.

## Limitations of the Current Standard

Looking at the scope of this proposal, the current DICOM Standard shows the following shortcomings:

* Missing definition for the efficient storage of stereoscopic reconstruction data from continuous video signals
* Missing specific solution for the structured mapping and storage of data from multiple intraoperative sensors
* DICOM provides some definitions like the Acquisition Context Module, the Optical Path Sequence or the US Region Calibration Module but they don’t meet our requirements.

We need a specific in-depth description of VL camera parameters and calibration containing e.g., focal length, principal point, radial and tangential distortion parameters, entry pupil in the calibration module.

## Description of Proposal

For the description of the relation of the Stereoscopic Videostreams we intend to make use of the Stereometric Relationship IOD (<http://dicom.nema.org/medical/dicom/current/output/chtml/part03/sect_A.43.html>)

To overcome the aforementioned limitations we plan two steps:

1. Introduction of a VL Image Calibration Module for the description of VL camera and calibration parameters
2. Definition of a suitable solution to store multi-modal data types e.g. calculated depth information or navigational sensor data.

**VL Image Calibration Module**

* Defining a module to support the pinhole camera model, which is widely used as a de-facto standard in the fields of computer vision, optical metrology, and photogrammetry. Important lens and image sensor parameters are, e.g., focal length, principal point, radial and tangential distortion parameters, entry pupil.
* Adding parameters that describe the quality of the reconstructed 3D data from video sequences, e.g., computing the measurement uncertainty from values such as pixel size, sensor size, and video resolution in combination with a measured working distance of the VL imaging system.

**Output data mapping to IODs**

The output data can be subdivided into three main classes as depicted in Figure 1 Systems Output Relation Diagram. Figure 2 Output data examples illustrates some examples of possible data belonging to the output data classes.

For a suitable representation there are now several possibilities that might work as solutions e.g.

* specific SR objects using the new table value type,
* new waveform objects similar to the Neurophysiology Waveforms,
* the proposed Whole Slide Microscopy Bulk Annotations Storage SOP Class.

Before a final decision is made this needs to be discussed within WG24 in conjunction with WG26. From the current status of the discussion within the working group for most of the data classes the work will most likely result in several modality specific new SR templates.

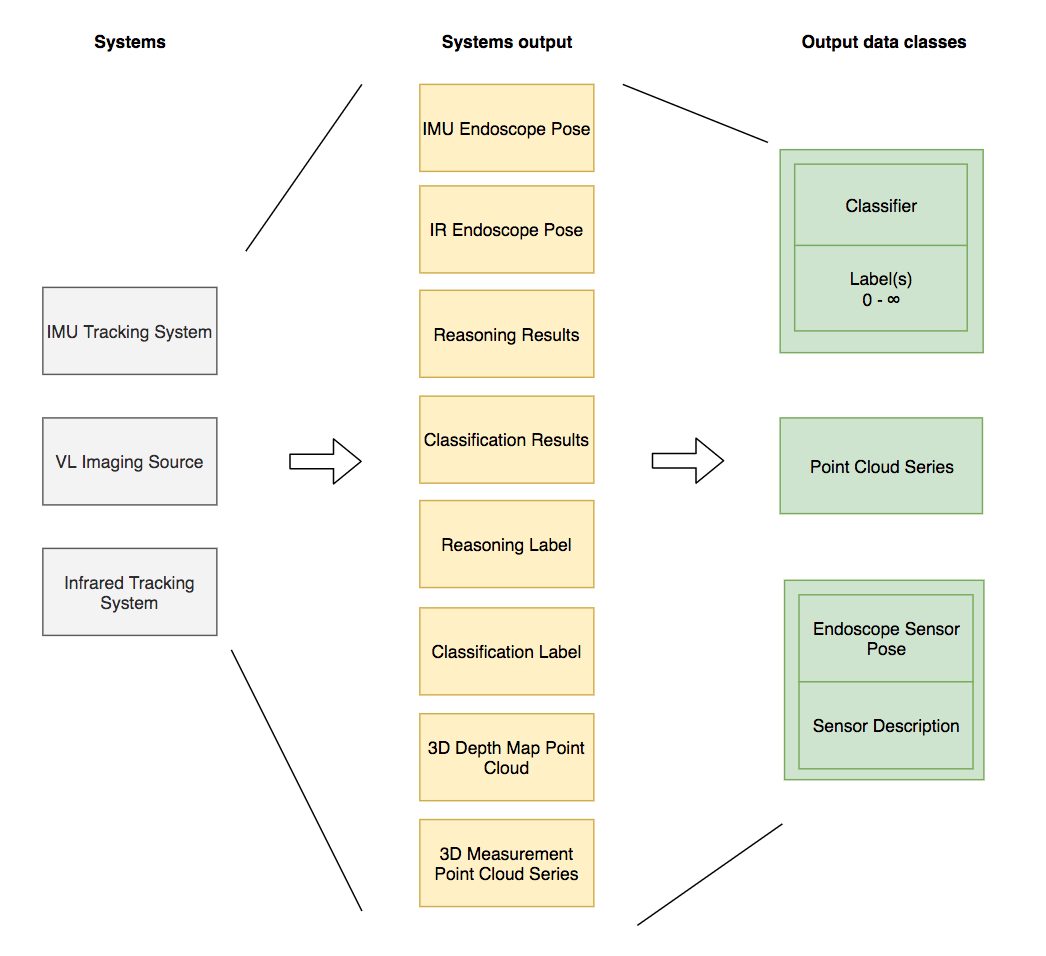
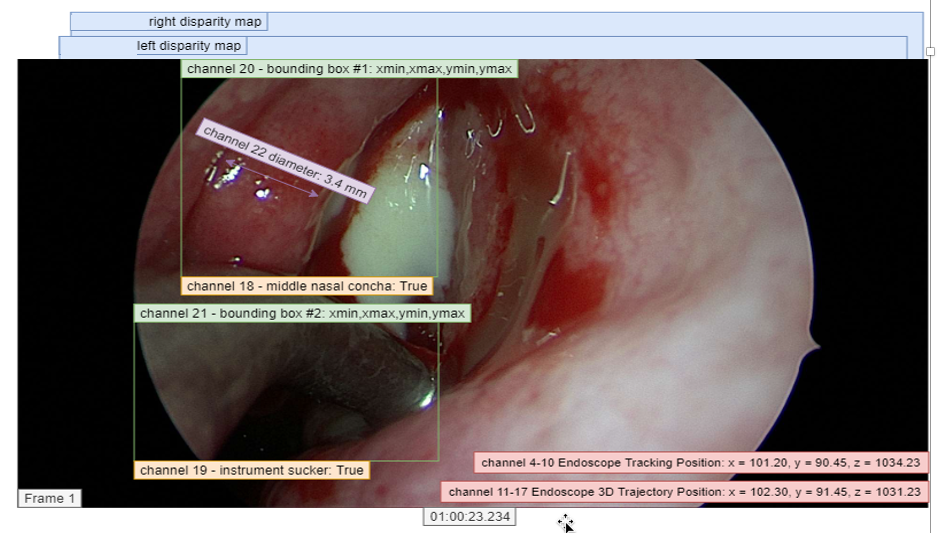


Figure 1 Systems Output Relation Diagram



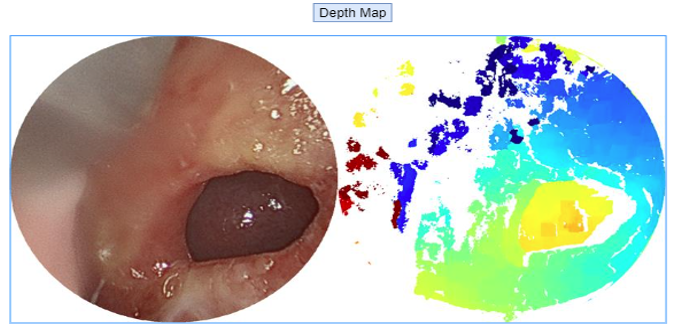


Figure 2 Output data examples

## Parts of Standard Affected

This work item will affect

* Part 3 – Adding additional calibration data to Video Endoscopic Image IOD
* Part 4 – If service objects have relevance for SR meta object
* Part 5 – Addition of various data references to Video Endoscopic Image IOD, new IODs
* Part 6 – If the addition of data dictionaries related to stereo-endoscopic parameters
* Part 16 – Content Mapping Resources
* Part 17 – adding use cases

## Resources & Time Line

About 10 people are active in WG-24. The main editors will be Marc Kämmerer (VISUS Health IT GmbH), Jean-Claude Rosenthal (Fraunhofer Institute for Telecommunications, Heinrich Hertz Institute, HHI, Berlin), and Richard Bieck (Innovation Center Computer Assisted Surgery (ICCAS), University Leipzig) of WG-24.

Additionally, the topic is addressed by the project consortium in the publicly funded research project “COMPASS” of the German Federal Ministry of Education and Research in the strategic innovation program “KMU Innovativ – Mensch-Technik-Interaktion” under the grant number 16SV8017.

A first draft worthy of consideration should be presented at the 2021 spring meeting of WG-06.

Our overall goal is to have a final proposal ready for the 2021 autumn meeting of WG-06 for balloting.

This topic is addressed by WG-24 and is endorsed by the following vendors:

* VISUS Health IT GmbH
* C.R.S. iiMotion GmbH
* Nuromedia GmbH
* Munich Surgical Imaging GmbH
* Schölly Fiberoptics GmbH
* Localite GmbH

Members of WG-24 anticipate that six to twelve hours of WG-06 meeting time will be required during 2021 to review and approve an early draft as well as a public comment, letter ballot, and final text versions of the supplement.