Title: Dynamic vision in robotics could bring big data to surgery

Authors: Mr Maxwell Flitton[1] and Miss Yasmin Jauhari[2]

Author information

1. Department of Medical Physics and Biomedical Engineering, University College London
2. Clinical Research Fellow, Clinical Effectiveness Unit, Royal College of Surgeons

Abstract

As big data is influencing more areas of medicine and surgery, the scalability, structure and management of data becomes more important. In surgical applications, data obtained from surgical procedures often require further processing and can be costly in storage. In this article, experimental dynamic vision computer technology and its science are introduced as a future option towards 3D mapping in robotic surgery.

Introduction

The term big data is trending; declaring its impact and cementing its vital role in a variety of industries[1]. This buzzword reflects not just the ability to gather large datasets, but also the opportunity to utilise machine learning to extract trends and produce predictive models. However, in this wave of exuberance, caution should be exercised against creating an environment where there is unsustainable interest that cannot be supported by the fundamentals such as tools, expertise and outcome.

Industry may have had a head start with its financial advantage and expertise to handle big data; but the authors propose that it is time for surgery to consider this investment, from the ground up. This is with the forethought of a different investment angle: aiming to reduce the overheads for storage and processing of data for future analysis. A major consideration with big data is the size of the data point and format of collection. As the volumes of data increase, so does the complexity of the hardware and expertise needed to process the data; effectively increasing the demand for time and resources. Therefore, in order for the big data “revolution” to truly take effect in surgery, smarter and cleaner data collection is required.

Even just focusing on the surgical operative setting, the potential utilities of collected data is overwhelming. One future impact area for big data is in 3D mapping, which is currently predominantly used in the pre-operative/ surgical planning setting. The aim of this article is to discuss 3D mapping in relation to big data, and potential use of computer vision towards the process.

There is a distinct difference between 3D mapping and 3D imaging. In 3D imaging, a 3D model of an object is created from a data source. Over the last few decades, this reformatting technology has developed alongside advances in computed tomography (CT) scanning and Magnetic Resonance Imaging (MRI)[2, 3], and can deliver details of lighting and texture[4]. This has been beneficial in the development of anatomical models for education[5] and in a range of surgical fields e.g. identification of vessels for perforator flap surgery, mapping complex cardiac structures for catheter ablation and navigation for neurosurgery[6, 7]. However, 3D imaging has limitations in relation to big data.

Although 3D imaging enables the surgeon to chart where they are in real-time, the technology is limited by its high-cost, hardware and processing power requirements and the need for preparatory complex scanning before and during the operation. These issues limit the scalability of 3D imaging to more procedures.

3D mapping in surgical big data

In contrast, 3D mapping calculates the camera’s position relative to the object in the field of view. Therefore, the whole section does not have to be imaged, just the areas of interest. A good illustration of the scalability of 3D mapping is self-driving cars.

Using a 3D imaging approach, the whole area would have to be scanned constantly before instructions can be fed to the car about its position. This would be a large burden on resources considering the hardware needed to track multiple cars in a large area and update these cars with their positions. Also, if the car wanted to go outside the area, it would be driving blind.

Using a 3D mapping, a self-driving car uses multiple cameras to detect obstacles, localise itself, and map the surroundings and path. Information about its position can also be integrated with information from its wheel speed and turning data, thus improving its accuracy[8]. The same scaling potential should to be considered for collecting data to map surgical procedures. 3D mapping would be able to evolve to collect increasing volumes of data on a range of surgical procedures. For example, 3D mapping the handle end of a laparoscopic surgical instrument with respect to reference points around the surgical hole can give the angle and depth of the surgical tool without the need for repeated and expensive scanning (Figure 1). This broadens the potential of data collection and mapping for even minor procedures.

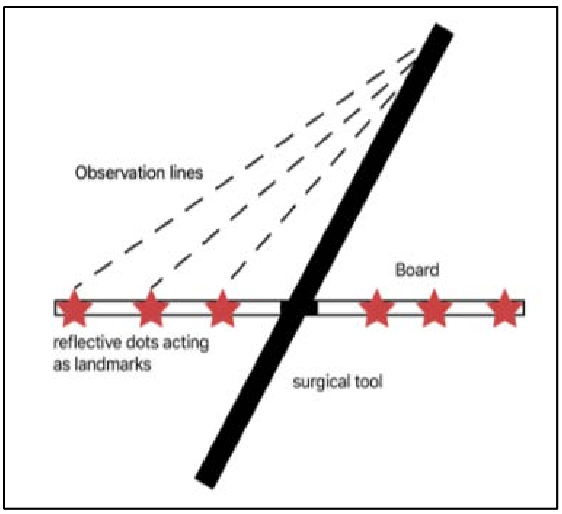

Figure 1: An illustration of how 3D mapping can be achieved by instrumenting the handle end of a laparoscopic tool.

Data collection in surgical big data

In minimally invasive surgeries such as robotics; 3D mapping is still a primitive technology[9]. Although there is a visual output in the form of a video feed, 3D mapping or analysing video feeds in this environment would consume a large amount of data.

Capturing video feed data from an individual operation using a regular camera would result in multiple data points (frames) per second, whereby each of these frames has a value for each pixel. This data (in video form) gathered for human interpretation for purposes of training, review or legal proceedings may be a wise trade-off for accuracy. However, in the context of big data, there would be a lot of redundant data collected which is not required, yet still needs to be stored in a database. Additionally, each frame would need additional processing, such as edge detection, to acquire further detailed information. This means passing a processing mask through every pixel in the image, and the work of this is multiplied for every frame. This exponential degree of processing must occur before a researcher can even consider modelling the operative data for an outcome, and would in totality require a lot of expensive hardware and expertise.

Dynamic Vision

With dynamic vision and advancements in computer vision, less expensive hardware and storage is needed to achieve this. As a dynamic vision sensor (DVS) does not store full frames it requires less storage. It also provides: 1. A timestamp, 2-3. A direction of movement for each point of interest (also known as an event, marked by X and Y coordinates) and 4. Information on light intensity. This output as a list of four numbers means that cheaper, commercial and less powerful computers can be used to capture, process and store data from fast moving dynamic frames using the DVS. Figure 1 illustrates how DVS events can be localises and time stamped in the experimental setting using a low power machine like a Microsoft Surface Pro[10]. This article will provide the science behind DVS, and the experimental procedures trialled to build a model for 3D mapping in surgery.

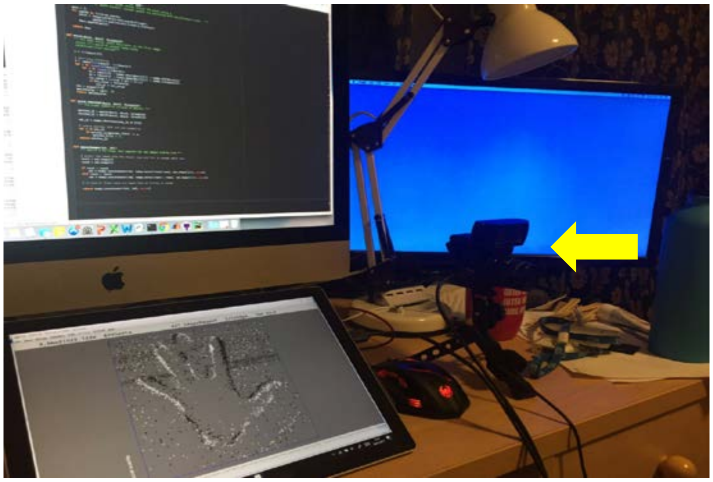

Figure 2: The use of a Dynamic Visual Sensor (arrow) with a commercial computer

The science of Dynamic Vision Sensors (DVS)

The way the DVS captures and sends data is a result of the difference in wiring. The DVS is inspired by brain architecture of neurons and is categorised as a neuromorphic electronic device. In the DVS, there are there are multiple sensors in the silicone retina that are asynchronous. When a sensor is activated (spike) for the pixel on the retina, the coordinates of the pixel and a timestamp is sent from the DVS to the computer. These motions captured by the DVS can be reconstructed by the timestamps, within a range of categories (bin sizes). Large bin sizes would group a number of timestamps together for one event. If the bin size was halved, you could assume that there would be double the amount of events This method of the asynchronous data capturing and reconstruction by timestamps means slow motion at very high frames (up to several KHz) can be achieved[11].

The DVS is also useful in low light environments. Normally, in this setting, a camera would have to enact exposure measures and loop them; with a consequent reduction to the frame rate. For example, if you were to wave a hand in front of a normal camera, without correcting the settings for low light, it would capture blurred motion. However, because of the increased frame rate offered, the DVS can have a larger dynamic range of illumination and process acceptable frame rates (of above 60Hz) in poor lighting[11]. This, combined with the reduced requirements for image processing, makes the DVS adaptable to rapidly moving, poorly lit closed, and cluttered environments[12].

Using DVS in 3D mapping

Advances in computer vision mean commercially produced cameras are able to calculate their position and 3D map the environment, as demonstrated by augmented reality technology on smartphones. This was made possible by the Zhang (2000) ‘pinhole camera model’ which enables the calculation of the camera’s position with relation to an object[13].

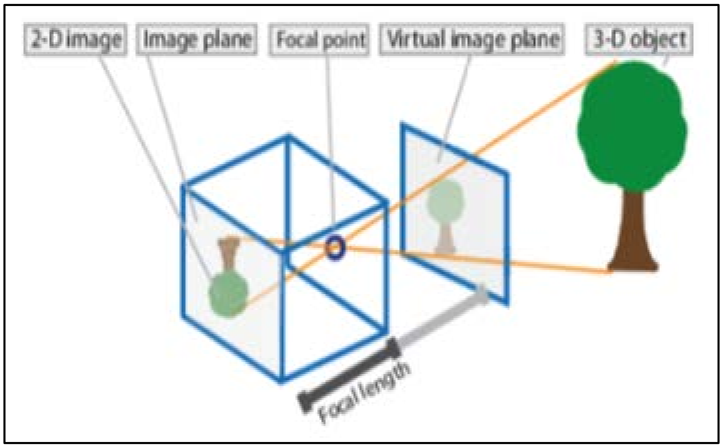

Figure 3: Pinhole camera model (15)
(Adapted from Mathworks (2017))

Inside a camera, a 2D plane of pixels can map a representation of the 3D world, whereby parameters such as the focal length of the lens can affect the 2D pixel representation ((camera) projection matrix). This means a camera model can be rearranged to calculate 3D coordinates from the 2D plane. As shown in figure 4, a pixel has X and Y coordinates (2D) in a camera image plane. This can be modelled by multiplying the object position in 3D with the camera matrix, which takes into account the camera's parameters. The result of this is a translated pixel position (figure 4).

$$\lambda \begin{bmatrix} x_c \\ y_c \\ w \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & c_{13} & c_{14} \\ c_{21} & c_{22} & c_{23} & c_{24} \\ c_{31} & c_{32} & c_{33} & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ T \end{bmatrix}$$

Position of the object in space (3D)

Camera parameters

Position of the Pixel on a 2D image plane

Figure 4: The annotated computational model for the pinhole camera to enable the calculation of an object position in 3D

With this in mind, in order to calculate an accurate image, a camera needs to be calibrated to its own parameters. A simple way of doing this can be by taking a series of photos of a flat plane such as a chess board at different angles. However, to calibrate the DVS, there needs to be a rate of change for the retina to pick up signals, whilst the device must remain still. One solution to this is a DVS calibration using a flashing grid of LEDs to solve the camera matrix[10]. This allows the calculation of 3D positions in space by multiplying the camera matrix by positions of the pixels in the image plane. Given that the DVS is an event based, low latency and data camera, the use of it with the pinhole model can provide a 3D mapping device that can capture data with low storage and processing needs.

Potential application of DVS in 3D mapping

There are several complex processes that need to be undertaken after calibration for 3D mapping to occur. At present, there are a range of online modules that can be downloaded and imported into a computer program. One example is ‘OpenCV’ - a free open source program which has built in calibration and 3D mapping functions[14]. Availability of complex programming on open source might diminish the need for individual expertise. This arguably takes each individual surgeon that one step closer to 3D mapping by attaching a DVS to a laparoscopic tool, and using a flat board as an operationalisation of the ‘pinhole model’ (Figure 1 and 3).

Limitations

It has to be noted that dynamic vision is new technology, and with that there are several limitations regarding the hardware and software that needs to be considered. Open software libraries can perform the task of calibration and mapping, but they do so with whole frames. This means data collected from the DVS will require formatting prior to software application. With regards to hardware, the DVS sensor would need to be adjacent to the handle of a surgical tool, and the noise created by this proximity could affect accuracy. In addition to this, there would need to be a reconfiguration of computer infrastructure and database codes to create a new storage system for the data. In the event that these limitations of accuracy and technicalities are resolved, and there is investment in expertise to support such a system; dynamic vision has a strong potential to be a viable tool in helping surgical big data processing and storage.

Summary

The DVS is one option towards delivering low cost, big data to surgery. It has the advantage of streamlined data with less of a demand on hardware and expertise on processing. However, it is acknowledged that the DVS is also likely to be one of many developing technologies that can be can used towards big data in surgery. In the future of surgery, it is therefore pertinent that any technology is appraised for its ability and ease of integration into day to day clinical practice. This technology should also minimise demand on resources yet yield information that will drive the big data revolution.

In summary, in order for surgical procedures to contribute to big data, it is important to wisely consider the method of data collection and resources needed to process these data points. Given the sea of opportunities for the utility of this data: training, research, personal development etc.; this is unlikely to be a futile investment.

References

1. Einav L, Levin J. Economics in the age of big data. Science. 2014;346(6210).
2. O'Connell RL, Stevens RJG, Harris PA, Rusby JE. Review of three-dimensional (3D) surface imaging for oncoplastic, reconstructive and aesthetic breast surgery. The Breast. 2015;24(4):331-42.
3. Salvolini L, Bichi Secchi E, Costarelli L, De Nicola M. Clinical applications of 2D and 3D CT imaging of the airways; a review. European Journal of Radiology. 2014;34(1):9-25.
4. McGhee J. 3-D visualization and animation technologies in anatomical imaging. Journal of Anatomy. 2010;216(2):264-70.
5. Bücking TM, Hill ER, Robertson JL, Maneas E, Plumb AA, Nikitichev DI. From medical imaging data to 3D printed anatomical models. PLoS ONE. 2017;12(5):e0178540.
6. Chae MP, Hunter-Smith DJ, Rozen WM. Comparative study of software techniques for 3D mapping of perforators in deep inferior epigastric artery perforator flap planning. Gland Surgery. 2016;5(2):99-106.
7. Rolf S, Hindricks G, Sommer P, Richter S, Arya A, Bollmann A, et al. Electroanatomical mapping of atrial fibrillation: Review of the current techniques and advances. Journal of Atrial Fibrillation. 2014;7(4):1140.
8. Häne C, Heng L, Lee GH, Fraundorfer F, Furgale P, Sattler T, et al. 3D visual perception for self-driving cars using a multi-camera system: Calibration, mapping, localization, and obstacle detection. Image and Vision Computing. 2017;68:14-27.
9. Overley SC, Cho SK, Mehta AI, Arnold PM. Navigation and Robotics in Spinal Surgery: Where Are We Now? Neurosurgery. 2017;80(3S):S86-S99.
10. Flitton M. Endoscope tracking in minimal invasive surgery using dynamic vision sensing: University College London; 2017.

11. Delbrück T, Linares-Barranco B, Culurciello E, Posch C, editors. Activity-driven, event-based vision sensors. Proceedings of 2010 IEEE International Symposium on Circuits and Systems; 2010 May 30 2010-June 2 2010.
12. Milford M, Kim H, Leutenegger S, Davison A. Towards Visual SLAM with event-based cameras. 2015.
13. Zhang Z. A flexible new technique for camera calibration. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2000;22(11):1330-4.
14. Laganiere R. OpenCV (2): Computer Vision Application Programming Cookbook. Birmingham: Packt Publishing Ltd; 2011.
15. Mathworks. What is camera calibration? https://uk.mathworks.com/help/vision/ug/camera-calibration.html?s_tid=gn_loc_drop2017 [Accessed 18/02/2018]