Constructing Three-Dimensional Models for Surgical Training Simulators

Marina Gavrilova[1], Stanislav Klimenko[2], Vladimir Pestrikov[2(✉)], and Arkadiy Chernetskiy[2]

[1] Department of Computer Science, University of Calgary, Calgary, Canada
[2] Institute of Computing for Physics and Technology, Moscow Institute of Physics and Technology, Moscow, Russia
pestrikov@phystech.edu

Abstract. New technologies introduced into medicine necessitate training of medical personnel to operate new equipment and techniques. For this purpose, training simulators and educational materials should be provided to the medical staff involved. This work concerns creating 3D model of surgical field for simulators to promote minimally invasive surgery. The paper reports the modes of constructing a photorealistic model of surgical field from endoscopic video streams, SFM, SLAM methods, as well as the problem of surface reconstruction from a point cloud and texture mapping on the constructed model.

Keywords: Surgical training simulators · Minimal invasive surgery · Structure from motion · SFM · Simultaneous localization and mapping · SLAM · 3D reconstruction

1 Introduction

In the modern world, it is evident that the new technologies appear practically every day. Being popular and spreading widely the recent technology advances easily integrate in our lives, with no exception to the medicine. New technologies available in medicine necessitate the medical personnel to be properly trained to operate with the new equipment and to handle new techniques. Therefore, it is necessary to create training simulators and provide educational materials. This paper concerns with the creation of 3D model of the surgical field for training simulators, training minimally invasive surgery.

The minimally invasive surgery is highly beneficial compared to traditional open operations. The whole procedure assumes small instrumentally-made cuts. The main advantages imply low blood loss, lower traumatism, and rapid and less painful post-operational recovery period. But such low-invasive intervention requires special training for surgeons.

Surgeons need to adapt handling a camera and operating instruments, to become accustomed to their mutual arrangement while observing an operation on a special monitor. These are the reasons to create modern surgical training simulators. A crucial function of simulators is to provide detailed and plausible imaging of the surgical field.

S. Siuly et al. (Eds.): HIS 2017, LNCS 10594, pp. 162–169, 2017.
https://doi.org/10.1007/978-3-319-69182-4_17

Thus, two major goals have to be reached: (1) to create qualitative and precise 3D model of surgical field, and (2) to map photorealistic textures on this model.

Firstly, the model should be properly constructed. There are two major approaches for constructing the operational field model: the construction may be done (1) by using the MRT data, medical atlases and other credible sources, the result of which can be exemplified by the ArthroS [1] simulator produced by EvenaMedical Company; or (2) by using photo- and video- images of real surgical operations, exemplified by scientific developments of A. Sourin [2–4].

In the second case, it is proposed to build a panoramic image of surgical field based on operation video. The relief model is then superposed on this panorama that a surgeon can interact with it and add three-dimensional models of target organs to the output. The drawbacks of this approach imply inaccuracy and distortions in transferring a voluminous object on a flat panorama, as well as a lack of possibility to utilize such model in simulator.

To avoid the above problems, it is suggested to begin with building of the 3D model of the surgical field using a video sequence. This procedure has got some advantages compared to constructing the model from the MRT data or biological atlases. It does not require preliminary location positioning, additional labels or an expensive preliminary examination (such as MRT or KT) as the implied method is based on information recovered from endoscopic images. In this case, the cost of creating each model is noticeably reduced. So, it is feasible to create the model of a particular object for its further detailed examination or for training and teaching purposes. For instance, a surgical field of appointed patient can be used as a source, the result of which might be helpful in modeling the course of a future operation. A surgeon can rehearse all complicated instances of a forthcoming operation, come across possible difficulties. In addition, it seems possible to create the models of particular clinical cases to be subsequently included into a scenario of personnel training.

For successful implementation, it is decided to use algorithms based on the Structure from Motion (SFM) and the Simultaneous Localization and Mapping (SLAM) methods. Noticeably, these methods, particularly SLAM, have been significantly developed recently.

The second part of this paper provides the basic steps of the SFM method. Then the paper introduces algorithms of the SLAM family. Further on, the modes of constructing 3D surfaces of the object and texture mapping onto the surface are described. The paper closes with the conclusion.

2 Principal Steps of SFM Method

The aim to construct 3D structures from a set of images (photographs or frames) taken from different perspectives is referred to as the structure from motion (SFM). It can be solved with the following procedure applied. At first, entry images should be analyzed to determine feature points. After that, coordinates of these points in three-dimensional space should be found using triangulation. In this respect the problem is to find the camera coordinates in the video source. With a sufficient amount of detected feature

points in the images, it is possible to compare the number of equations to define both coordinates of cameras and of feature points.

Thus, the steps for solving the SFM problem have been formulated [5]:

- **Search for features on input images.** In most cases it is advised to apply either SIFT or SURF for solving the problem of searching and describing feature points. SIFT commonly detects more features on the image, but SURF is more quick and seems to be more robust.
- **Search for matches between the points found.** It is recommended to refer to FLANN with the big high-dimensional data sets. It is also advised to use linear methods establishing the best compliance for each point found with sparse data [6].
- **The false compliances filtration.** Preliminary selection may be done basing on the assumption of positioning of corresponding points as well as on heuristic methods. These points are further filtered by the RANSAC method [7].
- **The equation system processing, the estimation of camera position and the reconstruction of the 3D structure.**

3 Algorithms of SLAM Family

The method of Simultaneous Localization and Mapping (SLAM) is currently widely applied. The problem posed for the method is classic for robotics: in transferring the sensor in the unknown space over the unknown trajectory it is advised to construct 3D or flat map of the environment and locate a sensor. In the case, when this sensor presented with a video camera, the SLAM problem and the SFM problem looks similar in nature. In this case it is called visual-SLAM or VSLAM. One of the differences to distinguish in the SLAM method is real-time working with a video, rather than with a final set of photos. That is, not the entire set of images is handed over to the algorithm as an input. Instead, the input data are portioned.

Keeping in mind the aforementioned, the working time is critical for SLAM-based algorithms, and it is reasonable to use the video sequence in their application. The key differences between the video stream and photos in respect to the paper issue are minor camera shifts on the sequenced frames, as well as the possibility of tracing feature points. Besides, the algorithm data is applied for the positioning autonomous devices and systems, which it is difficult or impossible to supplement with big computing power. Therefore, one of the main requirements for SLAM family algorithms is the demand of low computing power.

Today, two major procedures for solving Visual SLAM problem are outlined:

- *Feature-based methods* – the methods based on searching for feature points on the image;
- *Direct methods* – direct methods analyzing an entire image.

Each method is presented with its significant algorithm with the *ORB-SLAM* algorithm stands as a sample for feature-based methods and the *LSD-SLAM* algorithm – for the direct ones.

3.1 ORB-SLAM

The basics of the ORB-SLAM [8] algorithm are very similar to those of SFM. However, such features of SLAM as real-time operating and constant data updating amend the algorithm. Tasks of updating and specifying the model (or map) of the environment, constructing and refinement of the camera route, searching for loops on the camera route and some others appear consequently. In order to be equal to these tasks ORB-SLAM is designed to use the ORB algorithm [9] for features search and description. The algorithm is proved to have extremely high working speed and gives out the results of acceptable quality.

The authors of ORB-SLAM proposed their method for the environment map initialization. Firstly, two frames that have a sufficient amount of coinciding features detected with the ORB detector are picked. Then, basing on the matches found it is necessary to locate the camera shift in respect to selected frames. For cases when the camera is moved slightly, or part of scene getting into a frame is flat, the camera shift should be calculated via homography. Otherwise, the camera shift is better described with a fundamental matrix. The main idea of the method is that the initial camera position is computed with the two models involved. Then they are compared in order to look for the best one. The initial map of environment is built with the selected model applied.

ORB-SLAM consists of three modules that operate in parallel streams:

- Tracking stands for the module responsible for locating the camera position on the current frame, as well as for making the decision to distinguish a current frame as the key frame;
- Local mapping stands for the module responsible for construction and refinement of the model (or map) of the environment from distinguished key points;
- Loop closing stands for the module that searches for closing loop of a camera if a new key frame appears; it also refines the camera tracking and environment map if the loop closing is discovered.

As a result of applying this algorithm, the camera route is described with the decent precision and the point's cloud of the geometry of the environment is generated. However, the algorithm may interrupt its work or provide wrong results in case when a part of the environment contains small amount of features.

3.2 LSD-SLAM

The main feature of the LSD-SLAM algorithm [10] is that it uses the entire image as information pool, in contrast to other algorithms, which are feature point-oriented. Below the main stages of this algorithm are considered.

LSD-SLAM includes three main modules namely the tracking module, module of depth map estimation and the module of environment map optimization:

- The tracking module is responsible for locating the camera on each new frame. The camera position is calculated relative to the current key frame position. The position of previous frame is used as the initial approximation.

- The depth map estimation attracts the processed frame for refinement or replacement of the current key frame. If the shift between the new frame and the current key frame is small enough, then the depth map of the current key frame is specified. Otherwise, the new frame becomes a key frame. The depth map of the previous key frame is taken as the initial approximation of the depth map for the new key frame.
- When a new key frame appears, the previous key frame, which is not to be changed further, is used for updating the environment map.

The original depth map of the first frame is initialized at random. In the meantime the environment map construction is suspended for some time, until the algorithm output is reliable enough.

The noticeable advantage of the second method is the independence of feature points that may be extremely useful while processing poorly textured objects often found on the videos of minimally invasive surgeries.

4 Reconstruction of Surface and Superposition of Textures

The result map of the SFM approach and some SLAM algorithms, e.g. ORB-SLAM, is presented as a points cloud. But in a general case the result cloud is insufficiently rare for constructing a photorealistic model of a good quality.

4.1 Construction of a Dense Points Cloud

The most popular solution of the problem of constructing a dense points cloud is the method proposed in the referenced article [11]. This method is available in the PMVS library [12]. Each point from the rare cloud is superposed by a *patch* that is a small oriented rectangular. The same is done with neighbor areas of the model. Then incorrectly superposed patches are filtered. The last two steps have several iterations before the patches cover the surface of the reconstructed model densely.

Each image the algorithm receives as an input is conventionally divided into similar sectors $\boldsymbol{C_i(x, y)}$ in order approximate to the result in which every sector of each image contains the projection of at least one patch. The model of the patch $\boldsymbol{p}$ is described by the position of its center $\boldsymbol{c(p)}$ and a unit normal vector $\boldsymbol{n(p)}$.

A collection of images is picked for each patch showing the shared feature point. Some frames can be dismissed by the function of photometric error between the frames. Then parameters of a patch are specified via the minimization of photometric error between remaining frames. After that, when a corresponding patch is built for each point of the original points cloud, the algorithm comes over to the step of enlargement.

For each point of the image neighboring the cell that contains the patch $\boldsymbol{p}$ projection, the algorithm tries to build a new patch with the exception for the cells that already contains the projection of any patch. If the area of the model onto which a new patch $\boldsymbol{p'}$ is superposed has a significant height differences relatively to the area of the reference model containing the patch $\boldsymbol{p}$, the attempt to impose this patch onto the image reveals an easily noticeable error. In such cases the patch $\boldsymbol{p'}$ is set as an overshoot and is discarded from the model. If the patch undergoes the described test, its parameters are specified

via minimization of photometric error between the frames that contains the projection of this patch. After that, new patches are additionally filtered, and the process of constructing new patches is launched anew. The authors of the algorithm suggest repeating this procedure for at least three times.

Thus, in the algorithm output we get a new dense points cloud, after that it is possible to superpose meshes onto a scene and obtain the required surface.

4.2 Reconstruction of Model Surface from a Dense Points Cloud

After obtaining a dense points cloud, the surface model construction becomes accessible. This matter evolves a number of approaches, yet the authors of this paper consider the popular method of Poisson Surface Reconstruction [13] that available in the PMVS library most suitable.

The essence of this approach based on the observation that the vector field of internal normal of the solid body boundary may be interpreted as the membership function gradient of this body. The vector array may be taken as the vector field of the boundary. These vectors are inversed to patch normal obtained from the constructing of a dense cloud. Thus, the vector field $\vec{V} : R^3 \rightarrow R^3$ is obtained. It is necessary to find the scalar function $\chi : R^3 \rightarrow R\text{^}3$ minimizing the error function:

$$\chi = \int \left\| \nabla \chi(p) - \vec{V}(p) \right\|^2 dp \tag{1}$$

This problem is solving with the Poisson equation:

$$\Delta \chi = \nabla \cdot \vec{V} \tag{2}$$

Then the isosurface with a zero value of the function χ is found. The positive values of the function correspond to the original field of the model, while the negative ones - to external. All approaches and modifications improving the model are reported in the original paper.

4.3 Photorealistic Texture Mapping onto the Constructed Surface

Further on, it is necessary to map the textures onto the constructed 3D model. For this purpose the image of the due part of the model is projected on the surface of the constructed model from the positioned camera. The target part of the model is often found on more than one frame. So, it is critical to pick up the most appropriate frame for texture mapping. The following procedure is proposed for this particular instance. The points are estimated using several factors for each small part of 3D model. The list of these factors contains: (a) the distance from the model surface to the frame plane, (b) the angle between the frame and the surface normal, (c) the distance from the frame boundary (it is reasonable to map textures from the central part of a frame as optic distortions may occur on the periphery), (d) image definition, and (e) brightness (areas

that over lit or too dark are fined). Finally each part of the model is filled in with a texture from the frame with the highest score.

In this case, however, one more issue might be the case. Due to the difference in lighting, minor inaccuracies in the model construction and some misinterpretation of the camera location, the boundaries of the image textures stitching could be quite distinguishable. Thus, it is essential to level out the brightness of different frames and smooth out the transitions using the Multi-band Blending [14].

In this research, the texture mapping is an important issue to consider. The above-proposed algorithm combined with the correct choice of the 3D model reconstruction techniques, result in building the top-quality realistic model of surgery field that is the principal goal of this work.

5 Conclusion

The paper reports various approaches to constructing 3D model of environment from the video sequence or the collection of separate pictures. Application of such algorithms facilitates the creation of the surgical field model applied in surgical training simulators, and requires no specialist involvement. It stimulates building diverse models for medical personnel training, operative intervention planning or diagnostics.

Using the SFM procedure it is feasible to get a photorealistic 3D model of the environment, but in terms of the goal posed this approach has some disadvantages. For instance, this algorithm operates with a final set of static photos, rather than with a video stream. This means that before launching the algorithm it is necessary to distinguish an appropriate set of frames from a video sequence, and only after that one can proceeds with the 3D model construction. The set of random frames of a video sequence as well as the choice of frames with set intervals may be far from meeting the requirements. Besides, such procedure cannot operate in real time and build the model concurrently with the shooting. On the other hand, the presence of the entire set of entry data allows a coherent model to be built at once.

The paper reports the other approach, involving the SLAM application. It mitigates the issues of the SFM method described above, as they operate with a video stream in real time. With such procedure, the necessary key frames are selected automatically by the program, providing more optimal data set to build the 3D model with. An important advantage of this family of algorithms is their ability to work in real time with relatively low computing power. Thus, it can be concluded that SLAM is more suitable for solving the assigned problem.

The article reports two major approaches to copying with SLAM issues: (1) the method based on searching for feature points and (2) the direct method of analyzing the entire image. Comparing these approaches, the benefits of the feature points-based methods are: high-speed operation overshoots handling, essential stability and initialization easiness. The basic benefits of direct methods are: the ability of the algorithms to operate with poorly textured objects as well as with a few feature points on the images; the building of a more dense stage model, in contrast to the sparse points cloud of the aforesaid methods; applying the big piece of information obtained from the image.

Working with different objects under varying conditions, either approach should be preferred.

References

1. Stunt, J.J., Kerkhoffs, G.M.M.J., van Dijk, C.N., Tuijthof, G.J.M.: Validation of the ArthroS virtual reality simulator for arthroscopic skills. Knee Surgery Sports Traumatol. Arthroscopy **23**(11), 3436–3442 (2014). doi:10.1007/s00167-014-3101-7
2. Rasool, S., Sourin, A., Pestrikov, V., Kagda, F.: Modeling arthroscopic camera with haptic devices in image-based virtual environments. In: 2014 IEEE Haptics Symposium (HAPTICS), pp. 403–408. IEEE (2014). doi:10.1109/haptics.2014.6775489
3. Pestrikov, V., Sourin, A.: Towards making panoramic images in virtual arthroscopy. In: 2013 International Conference on Cyberworlds, pp. 48–51. IEEE (2013). doi:10.1109/cw.2013.29
4. Rasool, S., Sourin, A.: Image-driven virtual simulation of arthroscopy. Vis. Comput. **29**(5), 333–344 (2012). doi:10.1007/s00371-012-0736-6
5. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press, New York (2003). doi:10.1017/cbo9780511811685
6. Muja, M., Lowe, D.G.: Fast approximate nearest neighbors with automatic algorithm configuration. In: VISAPP 2009, pp. 331–340 (2009). doi:10.5220/0001787803310340
7. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM **24**(6), 381–395 (1981). doi:10.1145/358669.358692
8. Mur-Artal, R., Montiel, J.M.M., Tardos, J.D.: ORB-SLAM: a versatile and accurate monocular SLAM system. IEEE Trans. Robot. **31**(5), 1147–1163 (2015). doi:10.1109/tro.2015.2463671
9. Rublee, E., Rabaud, V., Konolige, K., Bradski, G.: ORB: an efficient alternative to SIFT or SURF. In: 2011 International Conference on Computer Vision, pp. 2564–2571. IEEE (2011). doi:10.1109/iccv.2011.6126544
10. Engel, J., Schöps, T., Cremers, D.: LSD-SLAM: large-scale direct monocular SLAM. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8690, pp. 834–849. Springer, Cham (2014). doi:10.1007/978-3-319-10605-2_54
11. Furukawa, Y., Ponce, J.: Accurate, dense, and robust multiview stereopsis. IEEE Trans. Pattern Anal. Mach. Intell. **32**(8), 1362–1376 (2010). doi:10.1109/tpami.2009.161
12. Furukawa, Y., Ponce, J.: Patch-based multi-view stereo software. http://www.di.ens.fr/pmvs. Accessed 10 June 2017
13. Kazhdan, M., Bolitho, M., Hoppe, H.: Poisson surface reconstruction. In: Symposium on Geometry Processing, pp. 61–70 (2006)
14. Burt, P.J., Adelson, E.H.: A multiresolution spline with application to image mosaics. ACM Trans. Graph. **2**(4), 217–236 (1983). doi:10.1145/245.247