

AUGMENTED 3D ENDOSCOPY VIDEO

A. Nedzved¹, V. Bucha², S. Ablameyko¹

¹United Institute of Informatics Problems of National Academy of Sciences of Belarus, Minsk, Belarus

²Samsung Research Center, Moscow, Russia

ABSTRACT

Estimation of 3D surface of a scene from a set of two or more images of that scene taken from the same point of view with different camera focal setting parameters is known as depth from focus/defocus problem. In this paper a system for 3D object reconstruction from endoscopy video is proposed which uses the same focal setting parameters for camera with different points of view of scene. In addition for preventing the undesired vessels dissection during operation, 3D video is augmented with extracted and recognized vessels. The vessels extraction is done with a novel real-time grey image thinning algorithm.

Index Terms— Virtual endoscopy, telepresence surgery, 3D reconstruction of medical images, shape from defocus, vessels extraction, grey image thinning

1. INTRODUCTION

Interactive systems that allow users to control and manipulate real-world objects within a remote real environment are known as teleoperator or telerobotic systems [1]. Such systems are often used in medical applications to confirm diagnosis and make telepresence surgery. The most known telerobotic surgery systems are da Vinci telerobotic surgical system [2] and ZEUSTM [3]. Such systems are controlled by surgeons remotely by viewing virtual surgical site with stereoscopic system and controlling stereoscopic camera and robot surgical armaments.

However such “stereoscopic” telepresence can be significantly extended and improved with augmented reality based on 3D reconstructed surface and extracted vessels. Such augmented 3D scene allow to:

- prevent undesired vessels dissection
- view the organ and disease from different points without camera movement;
- plan the strategy of surgery more precisely;
- teach the surgeons on the basis of obtained video and 3D reconstructed scene.

In this paper we focus on endoscopy surgery and methods for 3D surface reconstruction and vessels extraction. Vessels extraction and recognition algorithm is based on a novel grey thinning algorithm which is described in details further. 3D surface reconstruction from endoscopy

video sequence is based on depth from the focus/defocus approach [4, 5]. The image formation process can be described with optical geometry and the resulting image will appear focused and sharp only in a small 3D slice of the scene [5]. This fact is used in our approach to produce 3D surface reconstruction which consists from the following stages: image pre-processing, sharpness estimation for image features, estimation of features depth depending of sharpness, 3D slices generation with corresponding image features and depth, 3D reconstruction based on obtained slices.

2. DEPTH EXTRACTION FROM FOCUS\DEFOCUS

Endoscopy video is usually formed with a fiber-optic endoscope allowing physician to view patient’s esophagus, stomach and part of duodenum. For image forming a complex optical system is used. However, it can be approximated with aperture camera model.

The camera is always on the run and obtained image contrast depends on optical, acquisition and refining systems. Spatial geometrical and illumination distortions have a strong influence upon image quality also. Thus a pre-processing stage for image enhancement and distortion compensation is used for each video frame first of all and described in more details below.

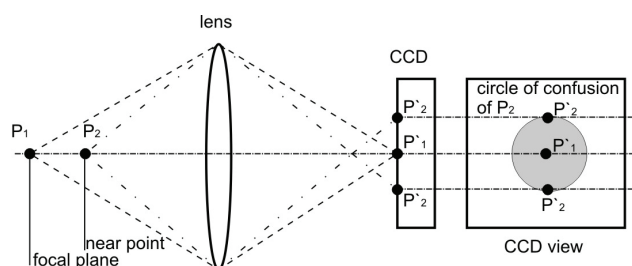


Fig. 1. Focus and defocus formation

According to the laws of geometric optics, the object point is represented on image by point, only if it is located at focal length from the lens (see point P₁ at Fig. 1). Any other object points which belong to non focal plane will be represented with circles of confusion (see point P₂ at Fig. 1). The larger diameter of the circle of confusion the unsharper image we get. As a result an endoscopy image *I* can be

treated as image which consists from sharp I_s and blur image I_b components.

The diameter of the circle of confusion depends on the distance to corresponding object point which can be estimated from the camera model.

The model of the camera can be approximated with the following well know equation:

$$1/f = 1/g + 1/b,$$

where g -distance to object, b -distance to CCD and f - a focus distance. The distance to sharp near and far points can be written as follows:

$$g = \frac{f^2 \cdot z}{f^2 \pm k \cdot z \cdot (g - f)}$$

where g - distance to object, f - a focus distance, k - aperture number, z - circle of confusion diameter.

The main idea of the depth extraction is to extract sharp image component I_s for the each frame from video sequence. In addition, we make an assumption that at each moment we exactly know the camera position and its movement step Δs . The analysis and depth estimation is applied for sharp image component I_s . The image I is transformed into slices according to depth. Each video frame during camera movement provides additional slices for volumetric representation which is used to reconstruct 3D endoscopy scene.

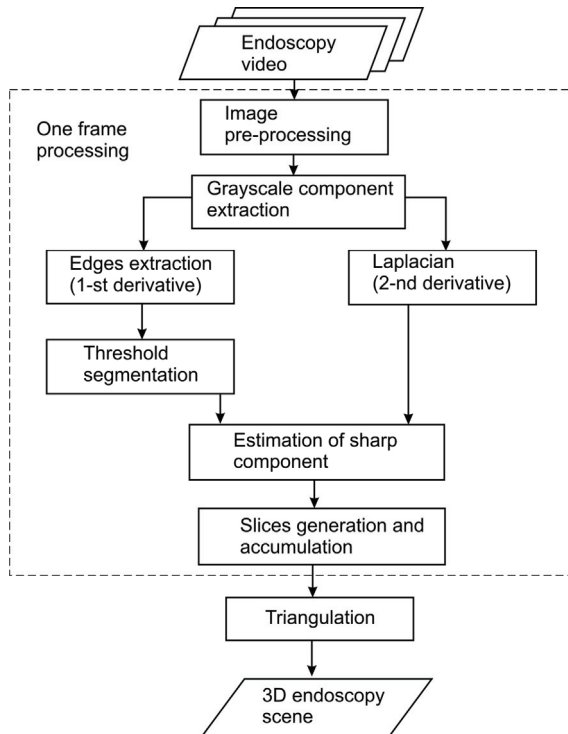


Fig. 2. Pipeline of the depth extraction

In general the pipeline of the proposed approach is represented at Fig. 2 which consists from the following main stages which are applied for the each video frame:

- pre-processing;
- sharp component extraction;
- slices generation;
- 3D reconstruction.

Sharp image I_s is represented with combination of two image features: contrast and edges. Such components are obtained with first and second derivatives processed with variation filter with kernel size 9x9 (See. Fig. 3, *b*). The result of sharp image extraction is a set of high-contrast curves (Fig. 3, *c*) which are obtained for each video frame.

As we know the camera parameters we can estimate the distance (depth) to high-contrast curves and obtain 3D points cloud which is used for 3D reconstruction. An example of depth generation by proposed approach is shown at Fig. 3, *d*.

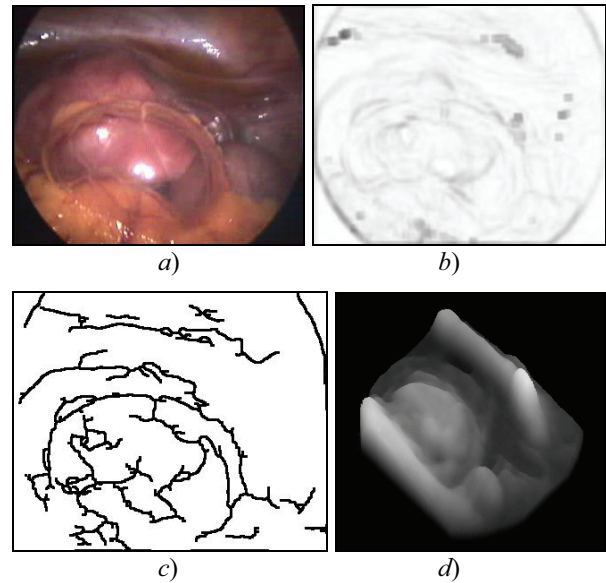


Fig. 3. Original image (*a*) and result of variation filtering of sharp component (*b*); high-contrast curves (*c*) and extracted depth from endoscopy video (*d*)

3. VESSELS EXTRACTION

Algorithm pipeline of vessels extraction is shown at Fig. 4 and corresponding steps are briefly described below.

Before vessels extraction it is necessary to improve image by pre-processing operation. This operation consists of two steps: noise removing and image enhancement. The noise removing depends from grabbing equipments and consists from a set of well-defined methods: low pass and median filtration, frequency analysis, etc.

The basic problem for vessel extraction is brightness distortions and background topology. In addition, geometric

distortions and shadows complicate the process of vessels recognition scientifically. Since object topology and most shadows have high geometric frequency they can be detected by frequency analysis. Hence topology background is determined by cutting of Fourier pattern of image.

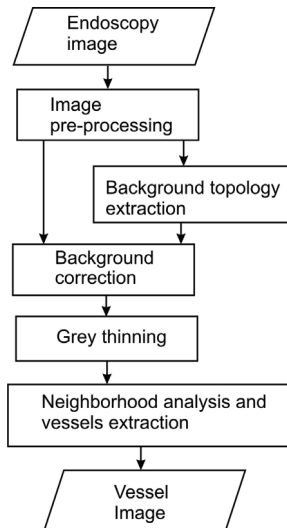


Fig. 4. Pipeline of the vessels extraction algorithm

For draft vessels extraction the enhanced source image and topology background are processed by background correction operation (See. Fig. 5, *b*).

The next problem is separating vessels from shadows and noise. We propose gray thinning algorithm and medial line definition to solve this issue. More details about grey thinning algorithm based on distance transform described further.

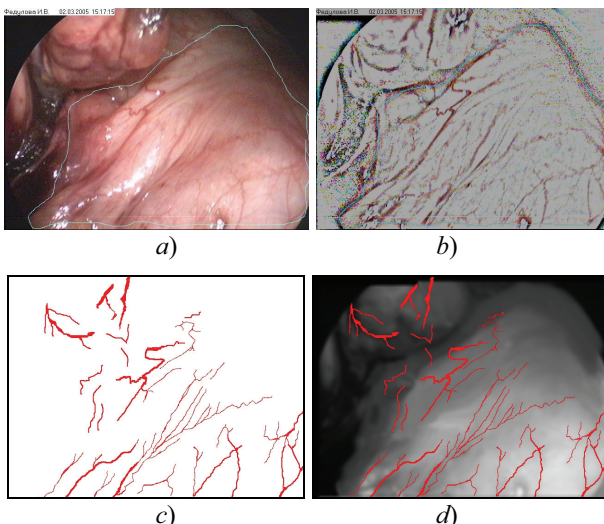


Fig. 5. Original image (*a*) and result of background correction (*b*); vessels extracted (*c*) and vessels overlaid on 3D model (*d*)

We use simple rules to recognize vessels after grey thinning operation which produces thinning image with medial thin lines. Objects with short medial line are noise and should be removed. For residuary objects classification local neighborhoods of medial lines are analyzed for gradient direction. If gradient is anisotropic then object is classified as shadow and removed. The residuary objects are classified as a vessels and form vessel image (Fig. 5, *c*) which is used as overlay image over initial video or 3D reconstructed model (Fig. 5, *d*).

4. GREY THINNING

Distance transform (DT) is defined as replica of the region of image where pixels are labeled with their distance from a reference pixels set of object [6]. For gray-scale image, reference set for such region is constituted by different gray-level. The union of DTs of regions of all gray-level is the DT of gray-scale image. When computing the DT of a region with gray-level k , adjacent regions with the levels greater than k are obstacles for the propagation of distance information (PDM).

For construction of a PDM, a gray-scale image is described as a collection of binary layers where every lower layer includes pixels from a higher layer. Distance map of one binary layer is constructed by increasing pixels depth for reflecting distance properties of pixels. Mostly gray-scale image has 256 levels. In this way, if every binary layer is raised to 256 levels, then we will have enough depth of gray value for construction of a distance map for every binary layer (Fig. 6).

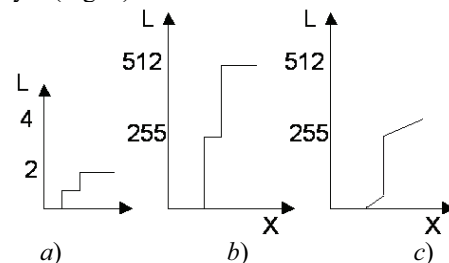


Fig. 6. Profile (L – pixels value, x – pixels coordinate) of (a) gray values in a gray-scale image (b) image with increased values, (c) a pseudo distance map.

On the base of obtained image with increased pixels value, distance map is built for every layer. The set of PDMs of all binary layers results to a pseudo-distance map with topological properties. In the result, we have a set of 256 image layers with 256 gray values for each layer.

We optimized a process of PDM building, and Chamfer metrics is used for optimal ratio velocity/quality. The PDM building algorithm contains two-scans. The first scan is realized in a direction from top to bottom and from left to right. For constructing PDM in this direction every pixel is changed by the following condition:

$$p = \begin{cases} (p_i + f_i), & \text{if } (p_i + f_i < p) \cup (p_i + f_i \geq L(p)), i = 0, \dots, n \\ L(p), & \text{if } (p_i + f_i < L(p)), i = 0, \dots, n \\ p, & \text{another case} \end{cases}$$

where: p is a pixel value, $L(p)$ – level of binary layer in image, p_i – value of pixels from a neighborhood, f_i – vector from mask-table for Chamfer metrics, i – index of element in mask table, n - number of elements in mask-table.

The second scan is realized by the similar condition with direction from bottom to top and from right to left and finalizes constructing PDM. This results to a pseudo-distance map with basic topological properties of an image. In this case PDM correspond to sets of layers of distance maps. In result pixel value of PDM include properties of gray-value and distance.

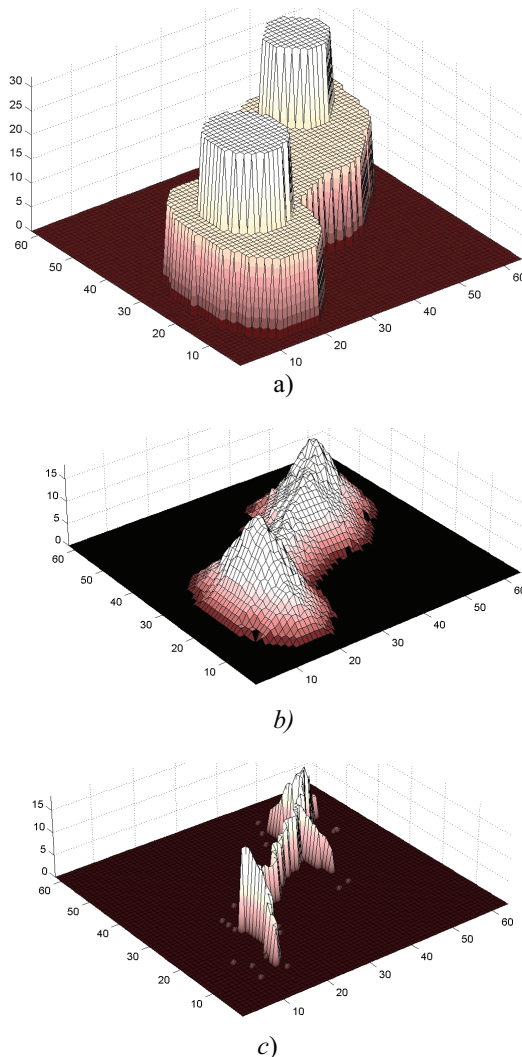


Fig. 7. An original object (a) and its pseudo-distance map (b) and grey skeleton (c)

Example of pseudo-distance map and grey level skeleton is shown at Fig.7.

Using the proposed pseudo-distance map, we can generally use any binary thinning algorithm that is oriented to the metrics that has been used for PDM building.

The algorithm for detection of skeleton pixels consists of two parts:

- 1) detection of feature pixels (topological properties):
 - a) end pixels;
 - b) saddle pixels;
 - c) duplex of saddle pixels;
 - d) local maximum.
- 2) Determination of remaining elements of a skeleton.

All pixels with previous topological properties are marked as a maximal level on the pseudo distance map. They are starting pixels for growing skeleton. After detecting starting point, function for finding of maximum is recursively executed in neighborhood of a pixel. This function ignores starting pixels. In a result of this operation, found maximum is marked as a next starting point. The skeleton grows until it does not reach other starting point or detects maximum.

5. CONCLUSION

Algorithm of vessels extraction from endoscopy video based on grey thinning concept is proposed. Extracted vessels are used to augment video and 3D scene of operation to prevent the undesired vessels dissection. An approach of 3D endoscopic scene reconstruction from endoscopic video is disclosed as well.

6. REFERENCES

- [1] O. Schreer, P. Kauff, and T. Sikora. *3D Videocommunication: Algorithms, concepts and real-time systems in human centred communications*, John Wiley and Sons, UK, 2005.
- [2] Ballantyne GH, Moll F. "The da Vinci telerobotic surgical system: the virtual operative field and telepresence surgery". *Surgical Clinics of North America*, Volume 83, Issue 6, pp. 1293-1304, 2003.
- [3] S.E. Butner, M. Ghodoussi, "A real-time system for tele-surgery". *Proceedings of International conference on distributed computing systems*, USA, pp. 236-243, 2001
- [4] S. Chaudhuri and A. Rajagopalan. *Depth from defocus: a real aperture imaging approach*, Springer Verlag, 1999.
- [5] P. Favaro and S. Soatto, "Learning Shape from Defocus," *Proc. European Conf. Computer Vision*, vol. 2, pp. 735-745, 2002
- [6] C. Arcelli, G. Sanniti di Baja and S. Svensson, "Computing and Analysing Convex Deficiencies to characterize 3D complex objects", *Image and Vision Computing* **23/2**, pp.203-211, 2005