21st Iranian Conference on Biomedical Engineering (ICBME 2014), Biomedical Engineering Faculty, Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran, Nov 26-28, 2014

Synthetic Data Generation for X-ray Imaging

Maryam Babaee, Ahmad Reza Naghsh Nilchi
Dept. of Computer Engineering, Faculty of Engineering,
University of Isfahan, Hezar Jerib Ave, Isfahan, Iran
{m.babaee,nilchi}@eng.ui.ac.ir

***Abstract*—Image processing softwares, like all softwares, need to be both verified and validated. Synthetic images are very useful during the medical software development process to verify the accuracy of algorithms. In this paper we introduce the process of generating synthetic 2D medical X-ray images in addition to ground truth imaging parameters. First, a 3D model of an organ (e.g., vessels) is made in a 3D-modeling software. Then, this volume model is voxelized based on the specified resolution in order to create a 3D CT image of that organ by assigning proper Hounsfield unit to each voxel. The obtained 3D CT image volume is used in DRR program as the input. Geometry parameters such as internal and external parameters are adjusted to take some images from different views. We demonstrated this process by three examples to confirm its usage in validation of medical image processing applications.**

Keywords*—*Synthetic Data, X-ray Imaging, Digitally Reconstructed Radiography

I. Introduction

Generating synthetic data is a necessary way to evaluate algorithms where the accuracy can be exactly verified under all conditions and consequently helps to gain a deep understanding of the algorithms behavior. For instance, it is useful to evaluate the degeneracy case of an algorithm, where in some situation, a specific algorithm might not work well. In the medical applications, synthetic data sets are much more important since the collection of medical images for research could be difficult due to the fact that medical images of individuals are private and normally hospitals and physicians are not allowed to distribute them. Additionally, data collection is difficult due to the use of harmful X-ray radiation, and contrast elements. These difficulties restrict the clinical data availability.

Synthetic data is a valuable tool, because the ground truth imaging parameters (i.e. both external and internal parameters of projection transformation) are available. For instance, we know where the images are taken and what the transformation between images is. Evidently, gaining these parameters from real data sets is not as accurate as synthetic data sets. Additionally, synthetic models can be used in designing a system and studying protocols and effective parameters related to the image acquisition.

Advances in processor architectures and computational systems have made computers able to solve difficult and complex problems. Nowadays, large number of images required for image analysis studies can be generated rapidly [1]. In this paper, we describe the process of generating synthetic images for X-ray based imaging modalities, like CT (Computed Tomography), Mammography and X-ray. The process starts first with 3D modeling of organs (e.g., vessels) and then voxelizing this model in order to create a 3D CT image by assigning proper Hounsfield units to each voxel. We utilize, a Digitally Reconstructed Radiography (DRR) [2] to capture the images from different views. The output of DRR is a number of images and also the internal and external imaging parameters. The obtained images can be used as ground truth data for further research such as 3D reconstruction [3], segmentation [4] and registration [5].

The rest of paper is followed as: In section II, some related works on generating synthetic X-ray images is discussed. In section III, we explain the process of synthetic data generation in more details. Here, we first talk about the image formation geometry in X-ray and explain the internal parameters as well as external. Section IV provides three practical examples of proposed method to generate synthetic images of vessels. Section V discusses about important factors which should be considered. Finally, in section VI we draw our conclusion.

II. Related Works

In generating synthetic data, a physical or digital phantom may be used. A phantom should have specific properties to mimic the behavior of a real data. For generating synthetic X-ray images of breast, a physical phantom could be prepared from mixtures of epoxy resin and iodine powder to behave like breast tissues [6]. Physical phantom also are applied for verifying medical simulations such as magnetic field tomography (MFT) inverse problem solutions [7]. A proposed method for production of synthetic X-ray images is applying the scale invariant feature transform (SIFT) in combination with epipolar morphing [8]. Monte Carlo techniques can be used in formation of direct X-ray images as well as exploring the effect of system parameters on image quality [9], [10].

III. Method

The proposed generating synthetic data in X-ray imaging is performed in several steps, which are explained in the fallowing sections.

A. 3D Modeling

First of all, it is essential to create the 3D model of the organ of interest, for example the vessel here, utilizing a 3D modeling software. For instance, we can use a commercial software like Autodesk 3DS Max [11] or an open source one like Blender [12]. 3D models can be created close to the real models by considering the radius varying along the vessel and the curvature of each branch. After designing the model, it is exported as an object file [13] for the next processing steps.

B. Voxelization

The vertices and faces of the 3D model saved in an object file are read by a Matlab toolbox, namely Toolbox Graph [14]. Then we consider a cube around the model and divide it to small pieces as voxels base on a specified resolution [15] .

Mammography is a specific kind of radiography in order to image the internal structures of breasts. Basic principal in CT and radiography are the same. In these imaging acquisitions, a X-ray beam is passed through the part of a body where makes X-rays absorbed or scattered. The remaining X-rays creating a pattern are transmitted to a detector (e.g., film or a computer screen) for more processing by a computer [16]. The amount of the attenuation of X-ray depends on the type of the substance placed in the path of the X-ray. Hounsfield Unit (HU) is X-ray attenuation unit used in CT scan interpretation. This unit determines the relative density of a substance. In other words, the point that how much of X-ray radiation is absorbed by each substance in tissue. Therefore, each voxel is assigned a value between -1000 (air, black) to +3000 (dense bone, white) as HU [17]. In a voxel with average linear attenuation coefficient μ_X, the corresponding HU value is therefore given by [18]:

$$HU = 1000 \times \frac{\mu_X - \mu_{water}}{\mu_{water}} \tag{1}$$

where, μ_{water} is the linear attenuation coefficient of water. Thus, a HU value according to the substance must be assigned to the each voxel of the volume. Next, this 3D array is written in a Meta-Image file, the so-called *MHA* format, as the data. This file consists of header information such as size of voxels, the dimension of image and etc. This Meta-Image file is the input file of DRR program to capture 2D images of it.

C. Generating X-ray Images

DRR is implemented as a module of Plastimath package [19]. Plastimatch is an open source software for medical image computation and processing. The Plastimatch has been developed based on ITK (Insight Segmentation and Registration Toolkit) [20] algorithms for medical image processing. The DRR program [2] takes a Meta-Image file as input, and generates multiple output images in format of *pgm*, *pfm*, or *raw* data. Additionally, the projection matrices and other geometric information such as intrinsic matrix, extrinsic matrix, and the center of image are saved in an ASCII file format. There are two type of calling DRR program; single image mode and rotational mode. In single mode, the geometry of the X-ray source and imaging panel must be determined completely. In the second case, the imaging geometry is supposed the default setting and some more options can be adjusted.

As it was mentioned, before generating images by DRR, it is essential to have a deep understanding about imaging geometry and how the mapping from 3D world to 2D image is done. A projection (camera) matrix is a 3×4 matrix which does this mapping. The camera matrix, in general, is the product of three matrices:

$$P = K[R|t] \tag{2}$$

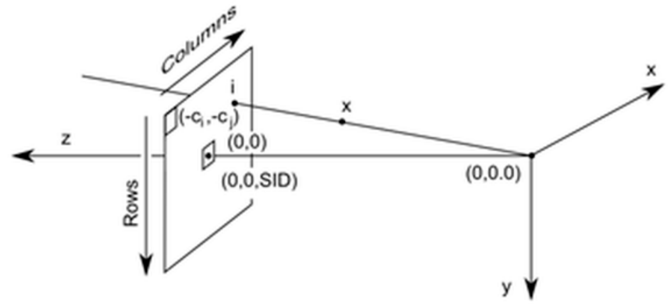


Fig. 1. World coordinate and image coordinate

Where R is a 3×3 rotation matrix and t is a 3-dimensional translation vector. When the first matrix is multiplied into the inhomogeneous representation of a 3D point and is added to translation vector, the result is the inhomogeneous representation of the point in the coordinate system of camera (i.e., here imaging system)

In DRR, P is a 3×4, the so-called *Projection Matrix* that maps homogenous world coordinates into homogenous pixel coordinates. A homogenous world coordinate is a 4 x 1 vector. A 3D coordinate (x, y, z) can be converted into homogenous coordinates by appending a 1: $(x, y, z, 1)$. On the other hand, a homogenous coordinate (x, y, z, w) can be converted into a 3D coordinate by first dividing each element by w, and then removing the last element: $(x/w, y/w, z/w)$.

If it is supposed that the world coordinate system is the imaging coordinate system and if the panel (i.e., image plane) is perpendicular to the z axis (Fig. 1) at a distance defined by the SID (Source to Image plane Distance), the mapping from world coordinate to image coordinate is done by *intrinsic matrix* K, where it is defined by

$$K = \begin{bmatrix} \frac{SID}{\alpha} & 0 & c_x \\ 0 & \frac{SID}{\beta} & c_y \\ 0 & 0 & 1 \end{bmatrix} \tag{3}$$

where α and β are the pixel spacing for columns and rows. Image center, (c_x, c_y) is the 2D pixel coordinate of the point on the image which is the intersection point of the line perpendicular to the image plane from X-ray source. In order to align the imaging system with the standard reference frame when the world coordinate system is different from the imaging coordinate system, it is needed to do a transformation defined by C, which is called *extrinsic matrix* and is given by

$$C = \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} \tag{4}$$

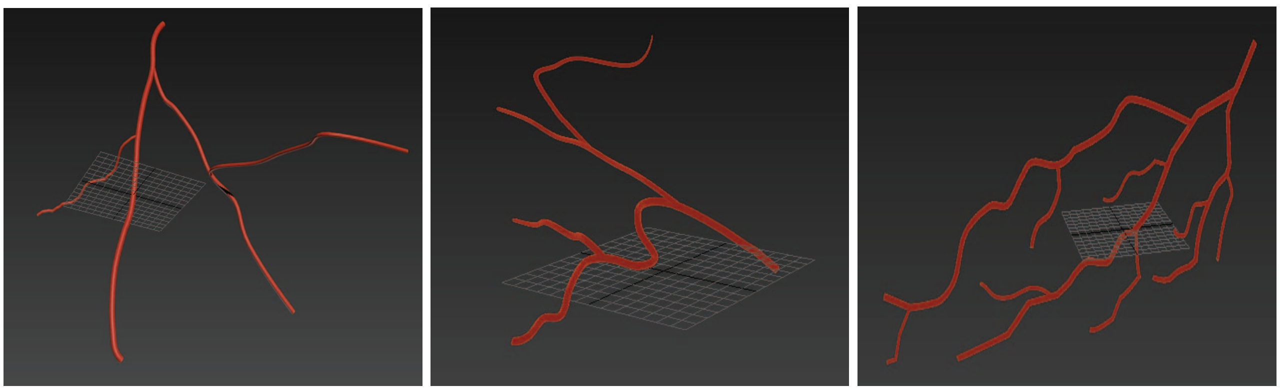
Where 0 is a 1×3 zero vector. The composition of the intrinsic and intrinsic matrices is called the *projection matrix*.

$$P = KC \tag{5}$$

Parameters used in calling drr command determine the geometry of imaging. Some of these parameters are SID, SOD (Source to Object Distance), resolution of the image and etc. Table I explains these parameters used in drr command.

TABLE I. ARGUMENTS OF DRR COMMAND [2]

Arguments	Explanation
-I infile	The input image (mha format)
-nrm ”xn yn zn”	The normal vector for the panel
-vup ”xv yv zv”	The vup vector (toward top row) for the image panel
-a number	The number of images generated from equally spaced angles
-N angle	Difference between two adjacent angles (in degrees)
-r ”row col”	Output image resolution (in pixels)
-c ”rowCenter colCenter”	The center of the image in pixels ; not necessarily (row/2, col/2)
-s scale	The scale of the pixels intensity in the output image
-g ”sad sid”	sad: source to object distance, sid: source to image panel distance (in mm)
-z ”s1 s2”	The physical size of imager (in mm)
-o ”o1 o2 o3”	Isocenter position
-A hardware	”cpu” or ”cuda” (default=cpu)
-t outputformat	Output file format: pgm, raw or pfm
-O outprefix	Determined destination for saving output files



(a) (b) (c)

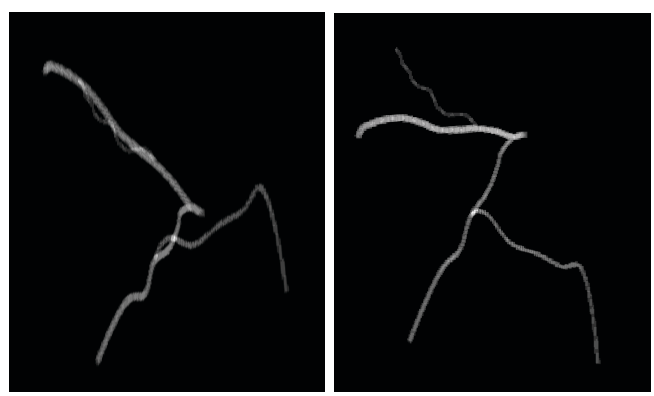
Fig. 2. Three vessel modeled in 3Ds Max software

IV. EXAMPLES

To show the examples of this method, some synthetic images generated from three vessel model (Fig. 2) are presented here. These synthetic data can be used in researches related to 3D reconstruction of vessels in mammography or angiography. Fig. 3 shows the drr command used in generating six images from third 3D vessel model. The difference between the neighboring angles is 30 degrees. The generated images from three models are shown in Fig. 4, Fig. 5, and Fig. 6, respectively.

```
drr -t pgm -s 15000 -a 30 -N 6 -g ”600 900” -o ”0 0 0” -z
”500 500” -r ”1500 1500” -c ”1450 700” -I model3.mha
-O model3
```

Fig. 3. drr command for generating six images from third model

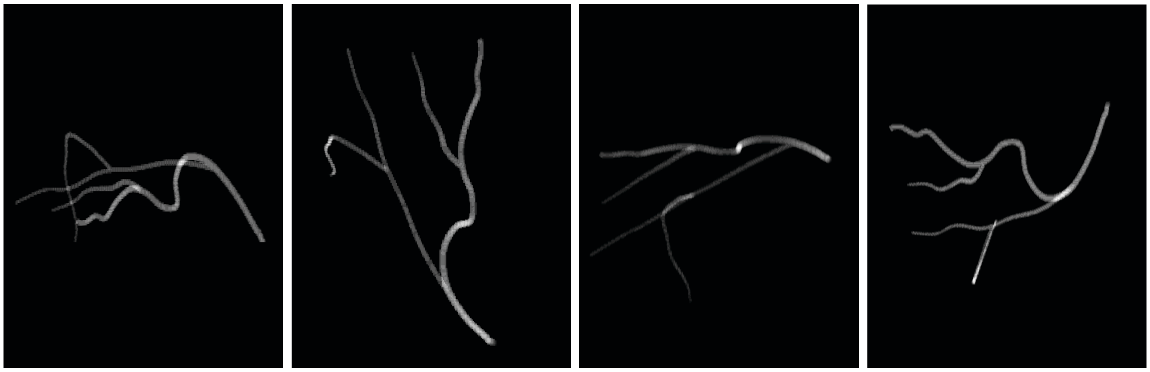


(a) (b)

Fig. 4. The two X-ray images generated from the first model by DRR

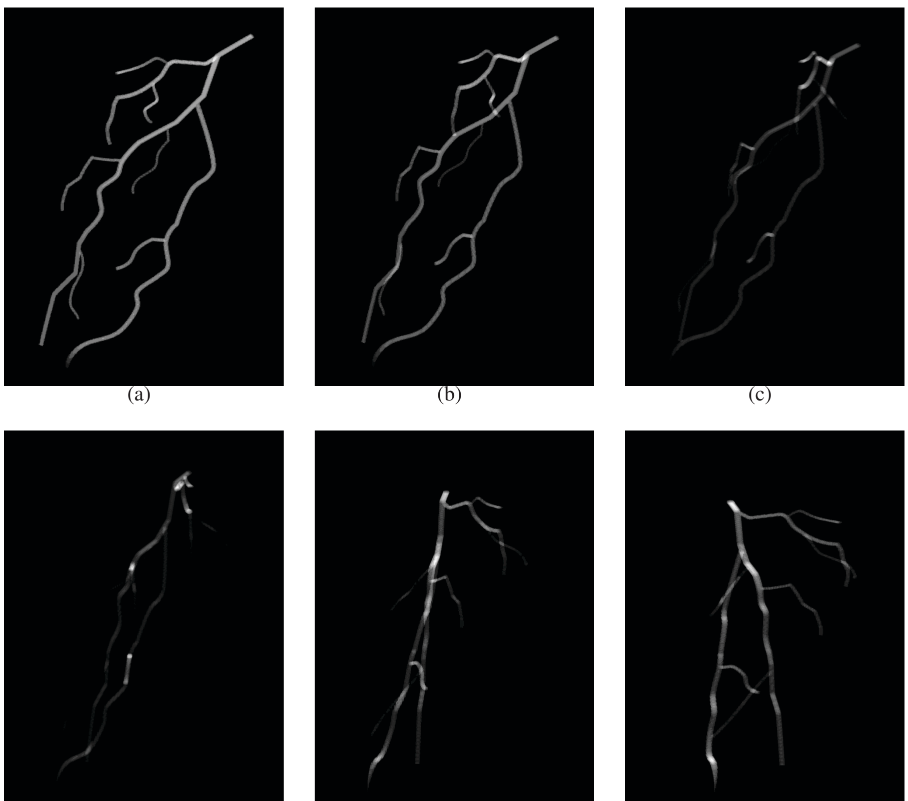
V. DISCUSSION

Due to existing difficulties in gathering the medical images, our introduced method can be useful in research academic projects. Obviously, the more similarity between the real mechanism and the simulation mechanism, the more similarity between the real and the synthetic images. To this end, it is recommended that selective parameters values be close the real ones. As we know, most image processing techniques work based on the intensity of pixels. Thus, one should pay attention to the values of some parameters such as resolution in voxelizing model, HU unit related to each voxel, and voxel spacing. These are some important factor in the intensity of pixels in the output image. One factor that effects on the intensity is the resolution of voxelizing the 3D model. Clearly, the less the value of resolution of voxelizing, the more number of generated output voxels. In this way, the number of voxels in the path of X-ray is increased. Therefore, the X-ray is weakened more and the intensity of pixel is higher. Of



(a) (b) (c) (d)

Fig. 5. The four X-ray images generated from the second model by DRR



(a) (b) (c)

(d) (e) (f)

Fig. 6. The four X-ray images generated from the second model by DRR

course, in addition of number of voxels, HU value assigned to voxels is an important factor in amount of this weakness. One disadvantage of this method is that the assigning HU value to voxels of a model consisted of several substance may be difficult. It is needed to determine each of voxels belongs to which substance. This step of algorithm and also voxelization of the model may be time-consuming. However, it is not a big issue because these step are not often performed during the research.

VI. Conclusion

We introduced a method for generating synthetic data in medical X-rag imaging. Since medical images with ground truth geometry imaging is hard to find, synthetic X-ray data are good option to evaluate an algorithm. In addition, these datasets will also be of great use for computer vision algorithms, because we provide the exact camera position and rotation in every frame. In this method, the 3D model of an object is created in a 3D-modeling software, at first. Afterwards, the 3D volume is divided to voxels based on a predefined resolution. After adjusting the geometry parameters of imaging, images are generated from the voxelized model by DRR as the way in X-ray imaging. This tool provides researchers ground truth data to validate their proposed method in medical image processing provided that they consider effective amount of each imaging parameter.

REFERENCES

[1] *Synthetic Medical Images*, 2012 (accessed September 14, 2014). [Online]. Available: http://www.cis.rit.edu/research/biomedical/projects/generatingSyntheticMedicalImages.html

[2] *DRR (Digitally Reconstructed Radiography)*, 2010 (accessed June 15, 2014). [Online]. Available: http://plastimatch.org/drr.html

[3] S. Habert, N. Dahdah, and F. Cheriet, "A novel method for an automatic 3d reconstruction of coronary arteries from angiographic images," in *Information Science, Signal Processing and their Applications (ISSPA), 2012 11th International Conference on*. IEEE, 2012, pp. 484–489.

[4] H. Prasantha, H. Shashidhara, K. Murthy, and L. Madhavi, "Medical image segmentation," *International Journal on Computer Science and Engineering*, vol. 2, no. 4, pp. 1209–1218, 2010.

[5] B. Zitova and J. Flusser, "Image registration methods: a survey," *Image and vision computing*, vol. 21, no. 11, pp. 977–1000, 2003.

[6] K. Bliznakova, I. Buliev, J. Kolev, Z. Bliznakov, and N. Kolev, "Study of suitability of new materials for use with physical breast phantoms," in *E-Health and Bioengineering Conference (EHB), 2013*. IEEE, 2013, pp. 1–4.

[7] H. Brauer, J. Haueisen, and M. Ziolkowski, "Verification of magnetic field tomography inverse problem solutions using physical phantoms," in *Computational Electromagnetics (CEM), 2006 6th International Conference on*. VDE, 2006, pp. 1–2.

[8] O. Abusaeeda, J. P. O. Evans, D. Downes, and J. Chan, "Animated depth employing image synthesis for 3d security x-ray imaging," in *Signal Processing and Information Technology (ISSPIT), 2011 IEEE International Symposium on*. IEEE, 2011, pp. 161–166.

[9] W. Zhao, W. Ji, and J. Rowlands, "Effects of characteristic x rays on the noise power spectra and detective quantum efficiency of photoconductive x-ray detectors," *Medical physics*, vol. 28, no. 10, pp. 2039–2049, 2001.

[10] N. Matos, "Monte carlo modeling of direct x-ray imaging systems," 2008.

[11] *Autodesk 3DS Max*. [Online]. Available: http://www.autodesk.com/products/autodesk-3ds-max/overview

[12] *Blender*. [Online]. Available: http://http://www.blender.org/

[13] *Object file format*. [Online]. Available: http://en.wikipedia.org/wiki/Wavefront_.obj_file

[14] *Toolbox Graph*. [Online]. Available: http://www.mathworks.de/matlabcentral/fileexchange/5355-toolbox-graph

[15] *3D Voxelizer*. [Online]. Available: http://www.mathworks.com/matlabcentral/fileexchange/21044-3d-voxelizer

[16] *Medical X-ray Imaging*, 2014 (accessed September 15, 2014). [Online]. Available: http://www.fda.gov/Radiation-EmittingProducts/RadiationEmittingProductsandProcedures/MedicalImaging/MedicalX-Rays/ucm2005915.htm

[17] *Hounsfield Unit*, 2014 (accessed September 15, 2014). [Online]. Available: http://www.syncopatient.com/Rad/CT/HnsfldUnt

[18] M. G. Kivelson and C. T. Russell, *Introduction to space physics*. Cambridge university press, 1995.

[19] G. Sharp, R. Li, J. Wolfgang, G. Chen, M. Peroni, M. Spadea, S. Mori, J. Zhang, J. Shackleford, and N. Kandasamy, "Plastimatch-an open source software suite for radiotherapy image processing," in *Proceedings of the XVIth International Conference on the use of Computers in Radiotherapy (ICCR), Amsterdam, Netherlands*, vol. 7, 2010.

[20] L. Ibanez, W. Schroeder, L. Ng, and J. Cates, "The itk software guide," 2003.