Detecting Mutual Natural Landmarks on MR and Camera Images

Ahmet E. Yetkin, Andac Hamamci
Department of Biomedical Engineering, Yeditepe University, Istanbul, TR
ahmet.yetkin@yeditepe.edu.tr, andac.hamamci@yeditepe.edu.tr

Abstract

In this study, landmark detectors which are trained on head model obtained using magnetic resonance (MR) images are applied on the optical camera images of the same subject. The main challenge in this problem is due to the domain discrepancy between the training and application domains. Hourglass alike CNN are employed to improve the domain generalization of the trained detectors.

1 Introduction

In image guided surgery, infering pose of body to register pre- and intra-interventional data is a well studied problem [1] . 2D-3D registration process includes merging 3D image modalities such as MRI, CT with 2D image modalities such as ultrasound, x-ray, optic camera image into the same coordinate system. In this study, we specifically address the problem of registering one's head MR volume to planar camera images, motivated by the incision planning in neurosurgery [2]. In [3], our approach is based on generating a head surface model using 3D MR volume and rendering a set of possible views under various pose and lighting conditions. By using this set, a convolutional neural network (CNN) is trained to directly estimate the rigid pose -hence the correspondence- on the camera images (Figure 1). This simple network resulted 95 percent success on images from its synthetic domain. However, the same system couldn't succeed on person's camera images due to the high domain discrepancy. Although the problem is too complex on global scale without the prior knowledge on the application domain, by employing decode-encode (hourglass) structures and intermediate supervision [4], the CNN classifier could be generalized to detect mutual landmarks on local patches from both domains [5]. It is possible to estimate the rigid transformation if sufficient number of such landmark correspondences could be reliably detected.

In this study, instead of training classifier on patches, a heatmap regression landmark detector [6] is used and its performance on detecting a sample natural facial landmark, which is labeled on MR volume, on camera images is evaluated.

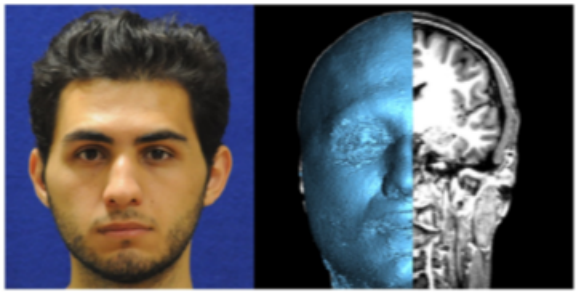

Figure 1: Images of person's camera image and 3D model obtained from his MRI.

1st Conference on Medical Imaging with Deep Learning (MIDL 2018), Amsterdam, The Netherlands.

2 Methods

2.1 Training Set

In this study, a healthy subject's head MR volume was acquired by a T1 weighted 3d spoiled gradient echo sequence in Yeditepe University Hospital. The mesh model for the subject's head is created by intensity tresholding followed by "Marching Cubes" triangulation in 3D Slicer software (*http://www.slicer.org*). The tip of the nose is selected as the landmark to be detected. Rendering operations were performed using VTK libraries and 3D Slicer's Python interactor. The head model is rotated with the angles sampled from uniformly distributed random variables $\theta_{YAW} \in (-90^o, 90^o), \theta_{PITCH} \in (-15^o, 15^o), \theta_{ROLL} \in (-15^o, 15^o)$. For each rendering a homogeneous light and directional light with a random direction and intensity is set using LightKit library [7]. 2100 synthetic views of the model is rendered (Figure 2) and the location of the nose tip in each view is tracked. For each view, a corresponding heatmap for nose tip is generated as a 2D isotropic gaussian function centered at the position of the nose tip and with a standart deviation of 1 pixel.

At the input of the network, each image is normalized to zero mean and unit variance. A uniformly distributed random noise in the range of the original dataset were added since the backgrounds of the synthetic set were completely uniform, which affect activations negatively.

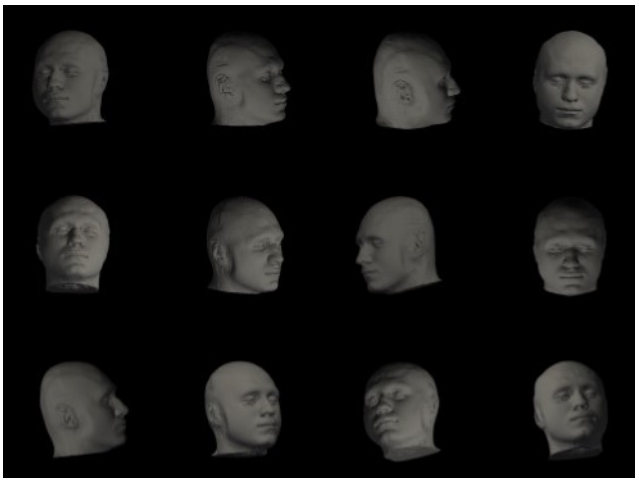

Figure 2: Samples from Training Dataset

2.2 Model Architecture

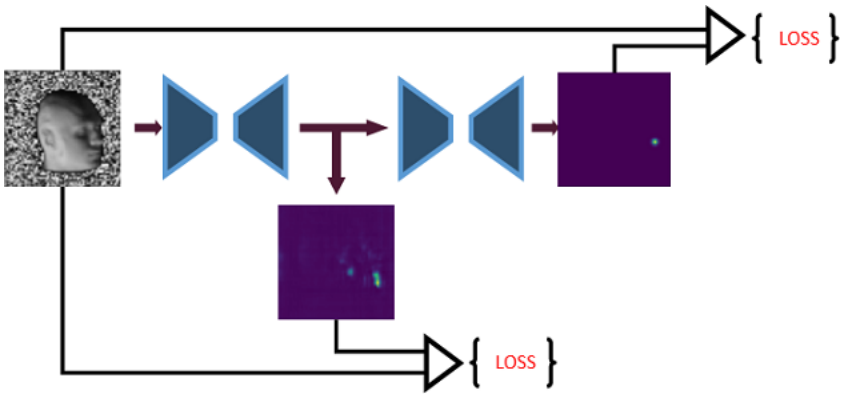

Figure 3: Model Architecture.

In this study, hourglass alike convolutional neural network given in Figure 3 was proposed [4]. It's main characteristic relates to autoencoder structures. It takes an input image, encodes by downsam-

pling with pooling and convolution layers. Therefore it represents its input by less parameters. Then it decodes with upsampling and convolutional layers, then produces heatmap by regression with same size of input image. This encoding-decoding process not only aims to understand images better, it also aims to make more invariant to size of objects that activates neurons. In this study, mergence of two hourglass structure was used by stacking them end to end. So the conv net model can be split into 2 sub models to be understood better. First model is single hourglass starts from input layer, ends with first heatmap generated. Second model starts with same layer that first model starts with. And contains both hourglass structures. First model was trained seperately. This intermediate supervision increases overall success of system. After reaching certain loss score, training of second model was initiated and weights of first model was used.

3 Results

Proposed model was tested on both synthetic images which are from the training domain and camera images which are from target domain. Results were categorized into 3 different classes correct, unique correct and negative. The sample images of the correct and uniquely correct images are given in figure 4. The correct and unique correct detection accuracies reported in Table 1 demonstrate similar correct detection performance for both synthetic and camera images.

Table 1: Test results obtained on both synthetic and camera domains

Test Dataset	Accuracy (Correct)	Unique Correct
Same Domain (Synthetic)	0.7	0.58
Target Domain (Camera)	0.8	0.2

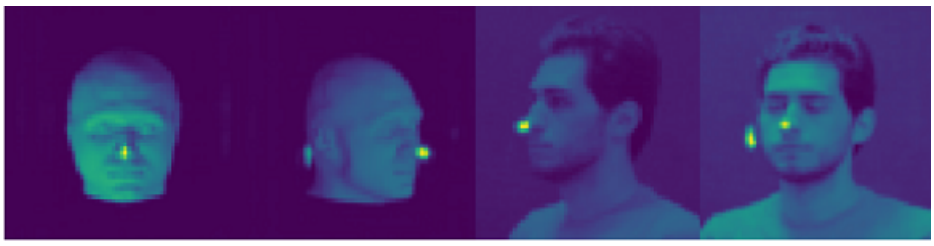

Figure 4: Sample results: (a) Unique correct on the synthetic domain, (b) Correct result on the same domain (c) Unique correct on target domain, (d) Correct result on target domain

4 References

[1] Markelj, Primoz et al. (2012) A review of 3D/2D registration methods for image-guided interventions, *Medical Image Analysis*

[2] Grimson, W. E. L., & Kikinis, R., & Jolesz, F. A., & Black, P. M. (1999). Image-guided surgery. *Scientific American*, 280(6), 62-69.

[3] Yetkin, Ahmet & Hamamci, Andac (2016) Data Augmentation for Head Pose Estimation from MRI Surface *TIPTEKNO Conference 16*

[4] NEWELL & Alejandro & YANG Kaiyu & DENG, Jia (2016) Stacked hourglass networks for human pose estimation. *European Conference on Computer Vision.* pp. 483-499

[5] Yetkin, Ahmet & Hamamci, Andac (2017) Region Proposal Networks in Domain Generalization of MR Landmark Detection *BIYOMUT Conference 17*

[6] Miao, Shun, & Z. Jane Wang, and Rui Liao. (2016) A CNN regression approach for real-time 2D/3D registration *IEEE transactions on medical imaging* 36.5: pp. 1352-1363.

[7] HALLE, Michael & MENG, Jeanette. (2003) LightKit: A lighting system for effective visualization. *Proceedings of the 14th IEEE Visualization (VIS'03) IEEE Computer Society*