National Repository
for Laryngeal Data

Removing Patient Information
from MRI and CT Images Using MATLAB

Eric J. Hunter [1,2,3,4] and Anil Kumar Reddy Palaparthi[1]

[1] National Center for Voice and Speech, The University of Utah
Salt Lake City, UT

[2] Department of Bioengineering, The University of Utah
Salt Lake City, UT

[3] Division of Otolaryngology – Head and Neck Surgery, University of Utah School of Medicine
Salt Lake City, UT

[4] Department of Communication Sciences and Disorders, College of Health, The University of Utah
Salt Lake City, UT

Abstract

Magnetic resonance images (MRIs) and X-ray computed tomography (CT) scans are essential tools in both the clinical and research environment for obtaining anatomical information from patients as well as from tissue samples. However, while the images themselves are not traceable to a person, digital information imbedded in the image files often contains personal information and, thus, may violate HIPPA laws or IRB protocol if the images were shared outside secure environments. This technical note gives a tutorial and downloadable computer scripts for quickly removing patient-specific information from folders containing many images.

U*pdates to this memo can be downloaded at http://www.nrld.org.*

Keywords: MRI, CT, patient information, vocal folds

1. Introduction

Magnetic Resonance Imaging (MRI) and X-ray computed tomography (CT) scans have become essential tools touching nearly all aspects of bioengineering research, particularly in biomechanical modeling (Hunter, Titze, & Alipour, 2004; Selbie, Zhang, Levine, & Ludlow, 1998; Titze & Hunter, 2007). The images can be used in obtaining anatomical information in clinical settings, human subject research and donated samples. These imaging techniques have advantages over traditional anatomical measurements (Hunter & Titze, 2005; Kim, Hunter, & Titze, 2004) because of their non-invasive nature. The imaging systems usually use DICOM (Digital Imaging and Communications in Medicine) format. Each image file has a file header which contains identifying markers such as patient name, patient number, and technician. While convenient, these markers bar the dissemination of these images, even if the image itself is not identifiable. As such, the image files are confidential information and may be subject to HIPPA regulations and/or Institutional Review Board oversight regulations. This information can be removed manually from the header of the file, or commercial software packages can be used to remove the personal information automatically.

The purpose of this report is to not only provide an additional method to removing personal information but a way to quickly remove the information in large folders of files; in other words, the purpose is to provide a set of MATLAB scripts which will *anonimize* the files. Using these scripts, patient information can be removed in a batch mode from hundreds of DICOM image files at a time. Then unidentifiable files can be shared in some instances, although it is important to check with your own Institutional Review Board or local HIPPA compliance officer for specific requirements.

2. Computer Scripts

The MATLAB script used to remove patient information (called `NRLDanonimizerA1.m`) uses the MATLAB function `dicomanon` (which removes confidential medical information from a DICOM file); `dicomanon` is part of the MATLAB Image Processing Toolbox. The current version of `NRLDanonimizerA1.m` was tested in Windows 7 Professional with MATLAB 7.11 and Image Processing Toolbox 7.1. The script can be downloaded with this document (both as a compressed .zip file or as a text .m file).

For basic usage, copy the `NRLDanonimizerA1.m` script into the folder with the MRI/CT images and have MATLAB execute the script in that folder. The DICOM images to anonimize should have a file extension of `.dcm`. The script goes through each file, removes patient information, making a copy of the original file with an extension of `.old`, leaving the new file in the original file's place. It is recommended that you make a copy of your directory with your original files then apply the `NRLDanonimizerA1.m` to the copied directory. In the new files, the confidential patient information is removed so to implement PS 3.15 / E.1 / Basic Application Level Confidentiality Profile (as implemented by the MATLAB function).

For the current setting, the script "modifies the confidential data and updates particular confidential data" (MATLAB help manual for diconan) yet preserves the Study/Series/Image hierarchy so that the order of the images is still known.

3. Examples and Variations

Deleting original files

After the `NRLDanonimizerA1.m` script has run, you may want to delete all of the original files from the directory, those with the `.old` extension. This can be done by typing the following at the MATLAB command line:

```
delete('*.old')
```

This command will delete all of the files with the `.old` extension.

Renaming Extensions

In some cases, the image files do not have a .dcm extension as required by the script, they could be .dicom or sometimes without an extension. In these cases, the files can be renamed in bulk and given an extension. For example, if the extension is .dicom, type the following command in the MATLAB command window:

```
!ren *.dicom *.dcm
```

Note that there is a space after '`ren`' and also after '`dicom`'. In the case of no extension, the following command can be used:

```
!ren *. *.dcm
```

Note that there is a space after 'ren' and before the second asterisks.

Errors in files

Occasionally there is an image file with a non-standard DICOM header. This can occur for several reasons but is rare. For example, here is the most common error:

```
:;problem with file D08332CD.dcm
{Warning: Attribute (0018,9305) has wrong number of values.}
```

For these files, the `NRLDanonimizerA1.m` script will add an 'x' to the beginning and end of the filename (e.g. `D082D1E5.dcm` becomes `xD082D1E5.dcm.x`). In addition, the filename of the offending file is output to the MATLAB command window and a summary text file named `anonymizer_results.txt` is created which will identify the problem files. Each file will have to be manually opened and patient information removed. While laborious, this can be done using a text editor like Notepad in Windows. Open the image and search for patient information.

Changing options of removed data

The script can be used to adjust different aspects of DICOM header information. To do this, lines 64, 69, 74 can be edited with `values` changed as needed.

```
dicomanon(d(p).name,sprintf([d(p).name(1:end-3) 'l'],p),'update',values);
```

4. More Information

For more information about DICOM images, conversions and applications, the reader is referred to the DICOM working group website (http://medical.nema.org/). Specifically, the website discusses general DICOM information but more importantly discusses the Basic Application Level Confidentiality Profile (http://medical.nema.org/dicom/2006/06_15pu.pdf).

References

Hunter EJ, & Titze IR. (2005). Individual subject laryngeal dimensions of multiple mammalian species for biomechanical models. *Ann Otol Rhinol Laryngol*, *114*(10), 809–818.

Hunter EJ, Titze IR, & Alipour F. (2004). A three-dimensional model of vocal fold abduction/adduction. *J Acoust Soc Am*, *115*(4), 1747. http://doi.org/10.1121/1.1652033

Kim MJ, Hunter EJ, & Titze IR. (2004). Comparison of human, canine, and ovine laryngeal dimensions. *Ann Otol Rhinol Laryngol*, *113*(1), 60–68.

Selbie WS, Zhang L, Levine WS, & Ludlow CL. (1998). Using joint geometry to determine the motion of the cricoarytenoid joint. *J Acoust Soc Am*, *103*(2), 1115–1127.

Titze IR, & Hunter EJ. (2007). A two-dimensional biomechanical model of vocal fold posturing. *J Acoust Soc Am*, *121*(4), 2254. http://doi.org/10.1121/1.2697573

Acknowledgements

Use Agreement

The scripts, images and text are open to use by the public as a service and part of the National Resource of Laryngeal Data (supported by the National Institute of Deafness and other Communicative Disorders). However, we ask the reader to respect the time and effort put into this manuscript and research.

If the text, images, or included scripts are used, the user agrees to reference this document, the NRLD, and the source of the original data. We also suggest that the user contact the original contributors of the data and give them the right of refusal to (1) participate on papers using the data and (2) have their supporting project acknowledged. The user agrees to freely share with the NLDR any extension software build on the data contained.

Revisions

1.0   Eric Hunter: Main document: original text (April 2011)
    1.1   Eric Hunter: additional DICOM description and clarification, minor edits, addition of Palaparthi as author (May 2011)
    1.2   Bob Jensen: minor edits (August 2011)
2.0   Laura Hunter: imported into new template, technical writing review (April 2015)