

Lossless Compression of Digital Mammograms

R. Visser, L. Oostveen, and N. Karssemeijer

National Expert and Training Centre for Breast Cancer Screening, Radboud University
Nijmegen Medical Centre, LRCB 577, P.O. Box 9101, 6500 HB Nijmegen, The Netherlands
`r.visser@lrcb.umcn.nl`

Abstract. The file size of images generated using digital mammography systems varies between 8 MB and 50 MB. The amount of data to be stored in digital screening programs is huge. Image compression may be helpful. In this study 8491 digital and digitised mammography images are compressed using 14 lossless compression schemes. The results show that using lossless image compression, the total amount of data to be stored can be reduced by a factor of 1.3 to 6.9 without loss of image quality. The actual data reduction depends strongly on the selected compression algorithm and the systems used to acquire and process the mammograms. The JPEG-LS and JPEG 2000 algorithms, both included in the DICOM standard, prove to be promising algorithms for screening programs because of the high compression ratios.

1 Background

For the design of the digital infrastructure for a population based breast cancer screening program, accurate data is required. The total amount of image data will have to be known in order to determine the feasibility of archiving and communication solutions.

The amount of image data can be reduced by applying image compression. Multiple studies on the compression of digitally stored medical images are known from literature [1,2,3,4]. From these studies it can be concluded that the achievable compression ratio depends strongly on the image type. In these studies however, a very limited number of digital mammography images was used. Therefore up to now there was no reliable information on the achievable compression ratios for digital mammograms.

The purpose of this study is to quantify the possibilities for reducing the required storage and network capacity by the usage of lossless image compression both for digital and for digitised mammograms. This data can then be used in the planning phase for the digitisation of screening programs for determining the required network capacity and storage space.

2 Method

2.1 Data Set

For digital mammograms the achievable compression ratio will depend on system properties like detector size and resolution, noise, homogeneity of the background and

Table 1. Number of images for each system and the specifications for the image types used

| System | Detector size (cm) ⁽¹⁾ | Pixel size (µm) | # Raw | # Proc | File size (MB) |
|------------------------|-----------------------------------|--------------------|--------------------|--------|--------------------|
| Agfa Embrace DM1000 | 17.9 x 23.3 | 70 | -(4,5) | -(4) | 16.3 |
| | 23.3 x 28.7 | | 88 ⁽⁵⁾ | 196 | 26.0 |
| Fuji Profect CS | 17.7 x 23.7 | 50 | 361 ⁽⁷⁾ | 361 | 32.0 |
| | 23.6 x 29.6 | | 40 ⁽⁷⁾ | 40 | 53.5 |
| GE Senograph 2000D | 19.1 x 22.9 ⁽²⁾ | 100 ⁽²⁾ | 1324 | 1363 | 8.4 |
| IMS Giotto Image MD | 16.3 x 22.8 | 81 | 542 | 543 | 10.8 |
| Kodak DirectView CR950 | 17.6 x 23.4 | 49 | 263 | 263 | 32.7 |
| Lorad Selenia | 17.9 x 23.3 | 70 | 234 ⁽⁵⁾ | 326 | 16.3 |
| | 23.3 x 28.7 | | 24 ⁽⁵⁾ | 36 | 26.0 |
| R2 ImageChecker | 15.7 x 23.7 ⁽³⁾ | 100 | 1212 | -(6) | 7.1 ⁽³⁾ |
| Sectra Microdose | 23.9 x 26.2 | 49 | 362 ⁽⁷⁾ | 357 | 49.7 |
| Siemens Novation DR | 17.9 x 23.3 | 70 | 6 ⁽⁵⁾ | 5 | 16.3 |
| | 23.3 x 28.7 | | 268 ⁽⁵⁾ | 277 | 26.0 |

4724 3767

¹ The *Imager Pixel Spacing* as specified in the DICOM header was used to determine the detector size.

² In the DICOM header of part of the GE-images the *Imager Pixel Spacing* was specified to be 94 micron, in the others it was specified to be 100 micron. The detector size was calculated using an *Imager Pixel Spacing* of 100 micron.

³ Because the scanned films were taken from an anonymised database, the patient labels had been cut from the images. Therefore the average image width was 15.7 cm in stead of the usual 18 cm. If the images had not been anonymised the file size would have been about 8 MB.

⁴ On the Agfa system at the BBNN only the full detector size is used. Therefore no images with the small detector size were available and the total amount of data stored is larger.

⁵ The Agfa, Lorad and Siemens systems use equal detectors. Therefore for the unprocessed images only differences in compression ratio's due to differences in the image sizes of the collected images are to be expected.

⁶ Although it would be good practise to optimise the digitised images for displaying, no processing was performed here.

⁷ The unprocessed Fuji and Sectra images are usually not available, but were made available by the manufacturers specifically for this project.