Similar Image Retrieval of Breast Masses on Ultrasonography Using Subjective Data and Multidimensional Scaling

Chisako Muramatsu[1(✉)], Tetsuya Takahashi[1], Takako Morita[2], Tokiko Endo[3,4], and Hiroshi Fujita[1]

[1] Department of Intelligent Image Information, Gifu University, Gifu, Japan
{chisa,tetsuya,fujita}@fjt.info.gifu-u.ac.jp
[2] Department of Breast Surgery, Nagoya Medical Center, Nagoya, Japan
takako@rose.sunnet.ne.jp
[3] Department of Advanced Diagnosis, Nagoya Medical Center, Nagoya, Japan
[4] Department of Breast Surgery, Higashi Nagoya National Hospital, Nagoya, Japan
endot@e-nagoya.hosp.go.jp

Abstract. Presentation of images similar to a new unknown lesion can be helpful in medical image diagnosis and treatment planning. We have been investigating a method to retrieve relevant images as a diagnostic reference for breast masses on mammograms and ultrasound images. For retrieval of visually similar images, subjective similarities for pairs of masses were determined by experienced radiologists, and objective similarity measures were computed by modeling the subjective similarity space using multidimensional scaling (MDS). In this study, we investigated the similarity measure for masses on breast ultrasound images based on MDS and an artificial neural network and examined its usefulness in image retrieval. For 666 pairs of masses, correlation coefficient between the average subjective similarities and the MDS-based similarity measure was 0.724. When one to five images were retrieved, average precision in selecting relevant images, i.e., pathology-matched images for benign/malignant index image, was 0.778, indicating the potential utility of the proposed MDS-based similarity measure.

Keywords: Image similarity · Image retrieval · Breast ultrasound · Breast masses · Mass classification · Multidimensional scaling

1 Introduction

Breast cancer is the most frequently diagnosed and one of the leading causes of cancer deaths for women in the United States, some European countries, and Japan [1–3]. Early detection and proper treatment can reduce the number of cancer deaths and improve patients' quality of lives. Mammography is considered the most effective method for screening breast cancers in general population with normal risk [4–6]. When an abnormality is found, additional image examinations, such as ultrasonography, are generally performed for diagnosis. Breast ultrasound is not only useful for differential diagnosis but also for screening in young women and/or women with dense breasts [7–10]. With

© Springer International Publishing Switzerland 2016
A. Tingberg et al. (Eds.): IWDM 2016, LNCS 9699, pp. 43–50, 2016.
DOI: 10.1007/978-3-319-41546-8_6

the approval of use of automated breast ultrasound system for screening by the Food and Drug Administration, ultrasound screening in adjunct to mammography is expected to increase. Therefore, computer-aided diagnosis system that can support an efficient multimodality reading may be useful.

We have been investigating a computer-aided diagnosis system that provides the similar images with known pathologies as a reference in the diagnosis of breast lesions on mammograms [11–14]. For retrieval of images that are diagnostically relevant and also visually similar, subjective similarity data by experienced radiologists for pairs of lesions on mammograms were obtained and used as the gold standard for training the system in our previous studies. The results of these studies indicated the potential usefulness of the method for determination of similarity measures using the subjective data. In general, masses with the same pathologies were considered more similar than those with different pathologies in terms of malignancy and benignity. However, it was more difficult to distinguish subcategories, such as cysts and fibroadenomas, on mammograms. It is expected that some lesions are more easily distinguished on ultrasound. Thus, an image retrieval system for ultrasound images can be useful. However, it is not known that our previous method can be effectively applied to ultrasound images. In this study, we investigated the method for determination of similarity measure for masses on breast ultrasound images and examined its usefulness in image retrieval.

2 Database

Breast ultrasound images used in this study was obtained at Nagoya Medical Center, Nagoya, Japan. This study was approved by the institutional review board. Ultrasonography examinations were performed as screening, follow-ups, or diagnostic work-ups using EUB-8500 (Hitachi Medical Corporation, Tokyo, Japan) or Aplio XG (Toshiba Medical Systems Corporation, Otawara, Japan) by diagnostic physicians or technologists together with physicians. When an abnormal lesion was found, two orthogonal views were generally captured. In some cases, two views were saved in a single image (Fig. 1(a)), whereas in others each view was saved in a single image (Fig. 1(b)). In this study, 14 images of benign masses, including 5 cysts, 6 fibroadenomas (FAs), and 3 benign phyllodes tumors, and 23 images of malignant masses, including 15 invasive

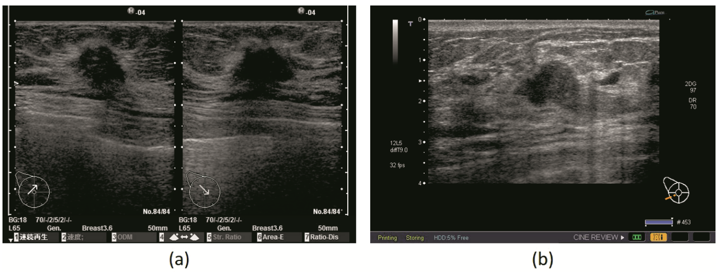

Fig. 1. Breast ultrasound images. (a) An image of orthogonal views of a scirrhous carcinoma (invasive ductal carcinoma) and (b) an image of a single view of a ductal carcinoma in situ

ductal carcinomas (IDCs) with 3 subtypes, 3 ductal carcinomas in situ (DCISs), 3 invasive lobular carcinomas, and 2 mucinous carcinomas, were selected as typical examples. They were employed in the observer study to establish the gold standard of subjective similarities of ultrasound masses.

3 Methods

3.1 Acquisition of Subjective Similarity Data

An observer reading study was performed for obtaining subjective similarity ratings for mass images. Using the 37 images, all possible pairs constitute 666 paired comparisons. Nine radiologists or breast surgeons who are certified for breast ultrasound reading by Japan Central Organization on Quality Assurance of Breast Cancer Screening participated in the study. Pairs of images were presented one by one as shown in Fig. 2, and each reader individually provided the similarity ratings on the continuous scale based on the overall assessment of shape, density and margin with respect to the anticipated diagnosis. The average ratings by the nine readers were considered as the gold standard (GS-A).

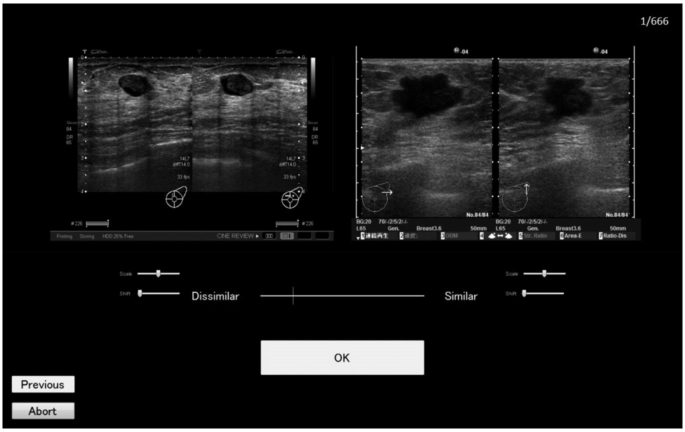

Fig. 2. User interface for obtaining subjective similarity ratings

Clinicians are not accustomed to evaluating similarities of lesions. Although some training cases were provided in the beginning, determination of similarity is not an easy task. In addition, reading of 666 pairs is a laborious task. There could be some pairs with large variations in their ratings. Thus, we also investigated the removal of outliers. For each pair, the ratings larger or smaller than 1.5 times the standard deviation from the average were excluded, and the average of the remaining ratings was considered as the alternative gold standard (GS-B).

3.2 Similarity Measure Based on Multidimensional Scaling

Multidimensional scaling (MDS) is a multivariate statistical technique which can display the dissimilarity relationship of data as a distance in lower dimensional space [15, 16]. It is a useful tool for visualization of the relationship with respect to the similarity between the subjects. In this study, the MDS was applied to the average dissimilarity (1- similarity) ratings for modeling the subjective similarity space reflecting the similarity relationship between the masses as illustrated in the top of Fig. 3. The number of dimensions was experimentally set to three in this study. Using the determined coordinates, a three-layered artificial neural network (ANN) with the back propagation algorithm was trained to estimate each dimension with image features as input data as illustrated in the bottom of Fig. 3.

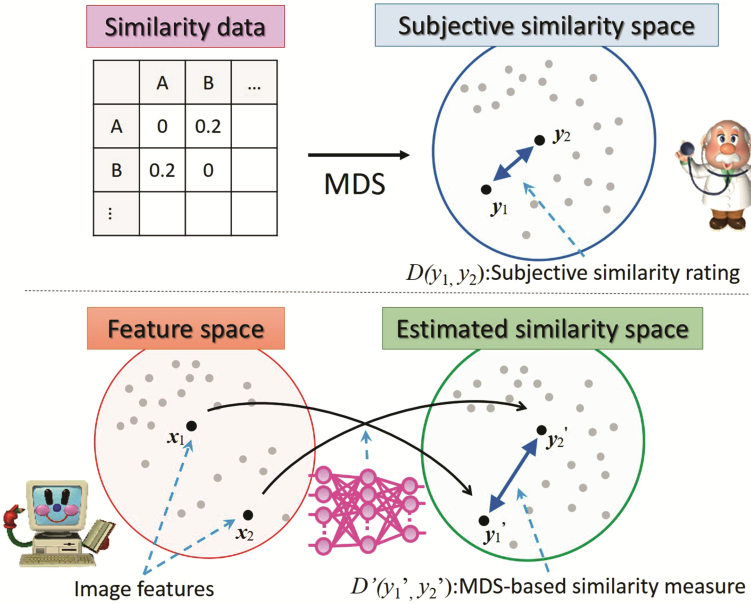

Fig. 3. Illustration of similarity space modeling

Eight image features characterizing the mass shape and four features related to the echo level were determined. The shape features included the ratio of height and width, circularity, irregularity, the degree of outline complexity, the number of dents in mass outline, the fraction of dented part in outline, the degree of dents, and the number of corners in outline. The height and width is determined as those of the rectangle enclosing the mass. The circularity is defined as the fraction of overlapped area of a mass and a circle placed at the center of the mass, whose area is equivalent to the mass. The irregularity is the ratio of the perimeter of the circle to the outline of the mass. The degree of outline complexity is the standard deviation of distances between points on the outline and the mass center. The number of dents is determined using the convex hull enclosing the mass. The degree of dents is defined as the ratio of areas of dents and the mass. The corners in outline are determined by sliding a circle on the outline and finding the area of overlap between the circle and the mass. Echo level features include the mean and standard deviation of pixel values, contrast, and posterior echo. The contrast is the difference of average pixel values in adjacent areas inside and outline the outline. The posterior echo is the ratio of average pixel values in the area posterior to the mass and the surrounding area. For the images with the combined two views such as in Fig. 1(a), the features were determined from both views and averaged. In the reconstructed space (MDS-based similarity space), MDS-based dissimilarity measure was determined by the Euclidean distance, which was then transformed to the similarity measure ranging from 0 to 1.

The MDS-based similarity measure was compared with the gold standard, and the usefulness of the measure for image retrieval was evaluated by the precision and recall. The precision is defined as the fraction of relevant images in the retrieved images whereas the recall is the fraction of relevant images that are retrieved to the total number of relevant images in the database. The images with the matched benignity or malignancy to an index image were considered relevant in this study. The similarity space modeling using MDS and ANN and the image retrieval test were performed by a leave-one-out cross validation test method. The performance was evaluated using GS-A and GS-B. For comparison, a conventional similarity measure based on the Euclidean distance (ED) in the feature space was also determined.

4 Result

Using all features, the subjective similarity space modelled by MDS was estimated by the ANN. The correlation coefficient between the GS-A and the MDS-based similarity measure was 0.724. When the GS-B was used in the space modelling, the correlation coefficient between the GS-B and the MDS-based similarity measure was 0.663. The correlation coefficient between the GS-A and ED-based measure was 0.514.

Using the MDS-based similarity measure, the most similar images for an index ("unknown") image were retrieved from the dataset (in this case from the 36 images). Figure 4 shows the precision and recall curves by using the GS-A, GS-B, ED-based measure, and the MDS-based measures using the GS-A and GS-B. The average precision at retrieving one to five most similar images was slightly higher when the GS-B was used than that using the GS-A (0.89 for the GS-B vs 0.88 for the GS-A).

When one to five most similar images were selected on the basis of the MDS-based similarity measure, the average precision was 0.78 and 0.81 using the GS-A and GS-B, respectively, for the training. They were not as high as those by the subjective ratings but higher than that using the ED-based similarity measure (0.70).

5 Discussion

It was expected that the subjective similarity ratings provided by experts were useful in selecting visually similar and diagnostically relevant images. Our previous study indicated the potential usefulness of the presentation of similar images in distinction between

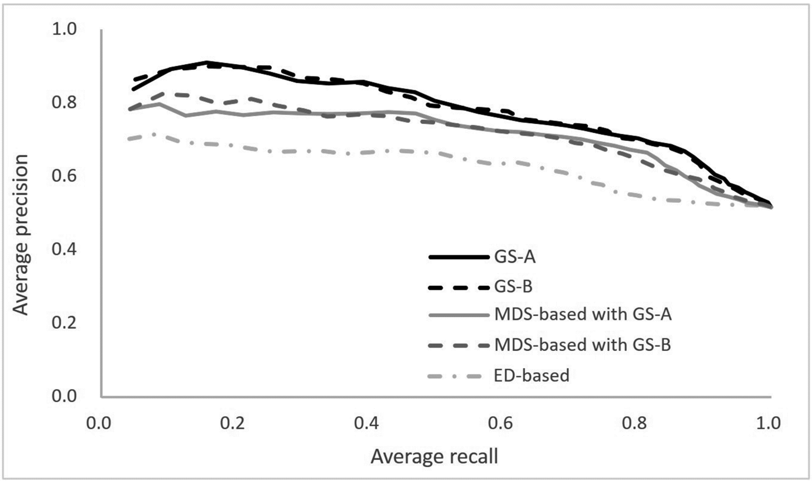

Fig. 4. Precision and recall curves for image retrieval using the gold standard and ED-based and MDS-based similarity measures

benign and malignant masses on mammograms [17, 18]. The high precision using the gold standard (GS-A and GS-B) for image retrieval in this study indicates the potential utility of the reference images selected on the basis of the image similarity by the experts for diagnosis of new lesions.

It was expected that the removal of outliers of subjective similarities may provide more reliable gold standard and improve the training process. The estimation of subjective rating using the ANN was more successful when the GS-A was used, which was suggested by the higher correlation. On the other hand, the average precision was slightly higher by using the GS-B when less than 5 images were retrieved, although it can be observed in Fig. 4 that the two curves are almost equivalent. These results may be due to the non-optimized parameters. Better parameter optimization process and the evaluation with an independent dataset are needed in the future. Also the usefulness of the presentation of images for the diagnosis of breast ultrasound images in clinical practice must be evaluated.

Overall, the result indicated that the MDS was effectively applied to model the subjective similarity space, and the similarity space was successfully estimated using ANN. Relatively high correlation coefficient between the gold standard and the MDS-based similarity measure and the high precision for image retrieval indicate the potential usefulness of the proposed similarity measure for selection of reference images that can be helpful in the diagnosis of breast masses on ultrasound images.

Acknowledgements. Authors are grateful to Mikinao Oiwa, MD, PhD and Misaki Shiraiwa, MD, PhD for their contribution in this study. This study was supported in part by the Grant-in-Aid for Scientific Research for Young Scientists (no. 26860399) by Japan Society for the Promotion of Science and Grant-in-Aid for Scientific Research on Innovative Areas (no. 26108005) by Ministry of Education, Culture, Sports, Sciences and Technology in Japan.

References

1. American Cancer Society: Cancer Facts & Figures 2016. American Cancer Society, Atlanta (2016)
2. Ferlay, J., Soerjomataram, I., Ervik, M., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D.M., Forman, D., Bray, F.: GLOBOCAN 2012 v1.0, Cancer incidence and mortality worldwide: In: IARC CancerBase No. 11 [Internet], International Agency for Research on Cancer, Lyon (2013). http://globocan.iarc.fr
3. National Cancer Center, Center for Cancer Control and Information Services: Monitoring of Cancer Incidence in Japan MCIJ 2011. National Cancer Center (2015)
4. Tabar, L., Fagerberg, G., Duffy, S.W., Day, N.E., Gad, A., Grontoft, O.: Update of the Swedish two-county program of mammographic screening for breast cancer. Radiol. Clin. North Am. **30**, 187–210 (1992)
5. Shapiro, S., Venet, W., Strax, P., Venet, L., Roeser, R.: Selection, follow-up, and analysis in the health insurance plan study: a randomized trial with breast cancer screening. J. Natl. Cancer Inst. Monogr. **67**, 65–74 (1985)
6. Humphrey, L.L., Helfand, M., Chan, B.K.S., Woolf, S.H.: Breast cancer screening: a summary of the evidence for the U.S. preventive services task force. Ann. Intern. Med. **137**, E-347–E-367 (2002)
7. Berg, W.A., Zhang, Z., Lehrer, D., Jong, R.A., Pisano, E.D., Barr, R.G., Bohm-Velez, M., Mahoney, M.C., Evans III, W.P., Larsen, L.H., Morton, M.J., Mendelson, E.B., Farria, D.M., Cormack, J.B., Marques, H.S., Adams, A., Yeh, N.M., Gabrielli, G.G.: ACRIN 6666 investigators: detection of breast cancer with addition of annual screening ultrasound or a single screening MRI to mammography in women with elevated breast cancer risk. JAMA **307**, 1394–1404 (2012)
8. Chae, E.Y., Kim, H.H., Cha, J.H., Shin, H.J., Kim, H.: Evaluation of screening whole-breast sonography as a supplemental tool in conjunction with mammography in women with dense breasts. J. Ultrasound Med. **32**, 1573–1578 (2013)
9. Brem, R.F., Lenihan, M.J., Lieberman, J., Torrente, J.: Screening breast ultrasound: past, present, and future. AJR **204**, 234–240 (2015)
10. Ohuchi, N., Suzuki, A., Sobue, T., Kawai, M., Yamamoto, S., Zhang, Y.F., Shiono, Y.N., Saito, H., Kuriyama, S., Tohno, E., Endo, T., Fukao, A., Tsuji, I., Yamaguchi, T., Ohashi, Y., Fukuda, M., Ishida, T.: J-START investigator groups: Sensitivity and specificity of mammography and adjunctive ultrasonography to screen for breast cancer in the Japan Strategic Anti-cancer Randomized Trial (J-START): a randomized controlled trial. Lancet **387**, 341–348 (2016)
11. Muramatsu, C., Li, Q., Suzuki, K., Schmidt, R.A., Shiraishi, J., Ewstead, G.M., Doi, K.: Investigation of psychophysical measure for evaluation of similar images for mammographic masses: preliminary result. Med. Phys. **32**, 2295–2304 (2005)
12. Muramatsu, C., Li, Q., Shiraishi, J., Doi, K.: Investigation of similarity measures for selection of similar images in the diagnosis of clustered microcalcifications on mammograms. Med. Phys. **35**, 5695–5702 (2008)
13. Muramatsu, C., Schmidts, R.A., Shiraishi, J., Li, Q., Doi, K.: Presentation of similar images as a reference for distinction between benign and malignant masses on mammograms: analysis of initial observer study. J. Digit. Imaging **23**, 592–602 (2010)
14. Muramatsu, C., Nishimura, K., Endo, T., Oiwa, M., Shiraiwa, M., Doi, K., Fujita, H.: Represenattion of lesion similarity by use of multidimensional scaling for breast masses on mammograms. J. Digit. Imaging **26**, 740–747 (2013)
15. Kruskal, J.B., Wish, M.: Multidimensional Scaling. Sage Publication, Beverly Hills (1978)

16. Shepard, R.N., Romney, A.K., Nerlove, S.B.: Multidimensional Scaling: Theory and Applications in the Behavioral Sciences. Seminar Press, New York (1972)
17. Muramatsu, C., Schmidt, R.A., Shiraishi, J., Endo, T., Fujita, H., Doi, K.: Usefulness of presentation of similar images in the diagnosis of breast masses on mammograms comparison of observer performances in Japan and the USA. Radiol. Phys. Technol. **6**, 70–77 (2013)
18. Muramatsu, C., Endo, T., Oiwa, M., Shiraiwa, M., Doi, K., Fujita, H.: Effect of reference image retrieval on breast mass classification performance: ROC analysis. In: Breast Image Analysis MICCAI, pp. 50–57 (2013)