Psychophysical similarity measure based on multi-dimensional scaling for retrieval of similar images of breast masses on mammograms

Kohei Nishimura*[a], Chisako Muramatsu[a], Mikinao Oiwa[b], Misaki Shiraiwa[b]
Tokiko Endo[b], Kunio Doi[c,d], Hiroshi Fujita[a]
[a]Department of Intelligent Image Information, Graduate School of Medicine, Gifu University, Japan
[b]Department of Radiology, National Hospital Organization, Nagoya Medical Center, Japan
[c]Gunma Prefectural Collage of Health Sciences, Japan
[d]Department of Radiology, The University of Chicago, Chicago, IL

ABSTRACT

For retrieving reference images which may be useful to radiologists in their diagnosis, it is necessary to determine a reliable similarity measure which would agree with radiologists' subjective impression. In this study, we propose a new similarity measure for retrieval of similar images, which may assist radiologists in the distinction between benign and malignant masses on mammograms, and investigated its usefulness. In our previous study, to take into account the subjective impression, the psychophysical similarity measure was determined by use of an artificial neural network (ANN), which was employed to learn the relationship between radiologists' subjective similarity ratings and image features. In this study, we propose a psychophysical similarity measure based on multi-dimensional scaling (MDS) in order to improve the accuracy in retrieval of similar images. Twenty-seven images of masses, 3 each from 9 different pathologic groups, were selected, and the subjective similarity ratings for all possible 351 pairs were determined by 8 expert physicians. MDS was applied using the average subjective ratings, and the relationship between each output axis and image features was modeled by the ANN. The MDS-based psychophysical measures were determined by the distance in the modeled space. With a leave-one-out test method, the conventional psychophysical similarity measure was moderately correlated with subjective similarity ratings (r=0.68), whereas the psychophysical measure based on MDS was highly correlated (r=0.81). The result indicates that a psychophysical similarity measure based on MDS would be useful in the retrieval of similar images.

Keywords: psychophysical similarity measure, multi-dimensional scaling, similar image, breast masses, mammograms

1. INTRODUCTION

In recent years, the number of women who are diagnosed with breast cancer is rapidly increasing and became the largest of all cancers in women in Japan [1]. Because the early stage cancers have good treatment outcomes and high survival rates, it is strongly desired to improve a response rate to screening mammography in order to detect cancer in early stage. However, in the screening, doctors must read a large volume of mammograms in a limited time, and therefore, it can be difficult to read them correctly without missing any abnormalities. For assisting image reading, computer-aided detection system has been introduced and employed in practice by many radiologists in the United States.

Once an abnormality is found, it is not easy to distinguish it between benign and malignant on mammograms. A number of computer-aided diagnosis (CAD) systems for distinction between benign and malignant lesions have been studied, and their usefulness has been demonstrated [2]. However, it may be difficult for radiologists to understand why a computer indicated a high (or low) number, when they use a CAD system with a numerical value of malignancy. Therefore, a CAD system which retrieves previous cases similar to an unknown case from a database and presents the medical information of the retrieved cases to the radiologists has been proposed [3]. Because such a CAD system may be adapted to clinical training, it is expected that the system must be easy for radiologists to utilize it.

For retrieving reference images which may be useful to radiologists in their diagnosis, it is necessary to determine a reliable similarity measure that would agree with radiologists' subjective impression. Similarity measures based on the distance in feature space is commonly used for retrieval of similar images. However, similarity measures based on the distance in feature space do not always correspond well with radiologists' subjective impression. In our previous study,

Medical Imaging 2013: Computer-Aided Diagnosis, edited by Carol L. Novak, Stephen Aylward, Proc. of SPIE
Vol. 8670, 86701R · © 2013 SPIE · CCC code: 1605-7422/13/$18 · doi: 10.1117/12.2001037

to take into account the subjective impression, the psychophysical similarity measure was determined by use of an artificial neural network (ANN), which was employed to learn the relationship between radiologists' subjective similarity ratings and image features. Similarity measures based on the distance in feature space was moderately correlated with subjective similarity ratings, whereas the psychophysical measure was highly correlated. The result indicated that the psychophysical similarity measure would be useful in the retrieval of similar images [3]. In this study, we propose a new psychophysical similarity measure based on multi-dimensional scaling (MDS) in order to improve the accuracy in retrieval of similar images, and demonstrated its usefulness.

2. METHODS

2.1 Image database

In this study, we used region of interest (ROI) images including breast masses on digital mammograms extracted by radiologists. These images were obtained by four different digital systems with different pixel sizes and gray scales. Therefore, in the preprocessing stage, the pixel size was normalized to 50 μm by linear interpolation, and intensity was down sampled to 10 bits. Masses with 9 pathologic types, including 3 benign types and 6 malignant types, were used in this study. The benign types include cyst, fibroadenoma, and phyllodes tumor, and the malignant types include ductal carcinoma in situ, papillo-tubular carcinoma, solid-tubular carcinoma, scirrhous carcinoma, invasive lobular carcinoma, and mucinous carcinoma. The malignant lesions were confirmed by biopsy or surgery, and the benign lesions were confirmed by biopsy or follow-ups by mammography and ultrasonography.

2.2 Acquisition of subjective similarity ratings

We performed an observer study for acquisition of subjective similarity ratings by 8 expert physicians, which were considered the "gold standard" of similarity measures. For the observer study, we selected 27 ROIs which is a total of 3 ROIs from each of 9 pathologic types. Each observer independently provided the subjective impression for all possible 351 pairs from 27 ROIs on a continuous rating scale between absolutely dissimilar (0.0) and absolutely similar (1.0), as shown in Fig. 1. They were asked to make the overall judgment based on shape, margin and density with consideration of expected pathology, however not to weigh on the size and surrounding normal tissue. The details of the method and the analysis of the results have been described elsewhere [4]. The average subjective similarity ratings were employed in the subsequent analysis.

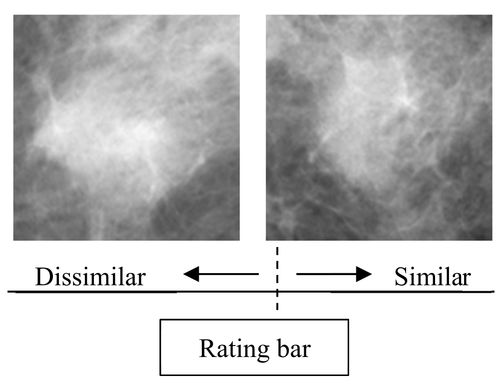

Figure 1. Method for determination of subjective similarity ratings

2.3 Extraction of image features

We determined 13 image features from each ROI image, which included 8 shape features, 2 density features, and 3 edge features. Shape features are the area, the diameter, the perimeter, the circularity, the irregularity, the ellipticity, the elliptical irregularity, and the minor-to-major axis ratio of a fitted ellipse. Density features are the contrast and the standard deviation of pixel values. Edge features are the average of edge gradient, the radial gradient index (RGI) [3], and the full width at half maximum (FWHM) of the radial gradient histogram [3]. The outlines of the masses were determined manually by a co-author (CM).

2.4 Determination of psychophysical similarity measure based on multi-dimensional scaling

MDS is a multivariate analysis that assigns locations to objects in the Euclidean space when provided the distance (or dissimilarity) matrix between them. As a result, it replaces similar objects nearby and dissimilar ones far apart. We determined coordinate values of each ROI in a multi-dimensional space (subjective space) by applying MDS to the radiologists' subjective similarity ratings. Because the subjective similarity rating has a qualitative characteristic, Kruskal's non-metric MDS [5] was employed in this study. In non-metric MDS, the stress, *S*, is defined as

$$S = \sqrt{\frac{\sum_{i<j}\left(d_{ij} - \hat{d}_{ij}\right)^2}{\sum_{i<j} d_{ij}^2}} \tag{1}$$

where d_{ij} represents the distance between i and j while $\hat{d}_{ij}$ represents the distance in the configured space, and the vector coordinates that minimize the stress will be searched. For minimization, a steepest descent method is used. Subsequently, an ANN was employed to learn the relationship between the coordinate values of each dimension in the multi-dimensional space and the image features. As a result, new coordinate values were determined with a combination of the image features by using the ANN, and a new similarity measure (i.e. psychophysical similarity measure based on MDS), which was scaled between absolutely dissimilar (0.0) and absolutely similar (1.0), was determined by converting the Euclidean distance between ROIs using an exponential function. For learning and testing, a leave-one-out method was used. The parameters of ANN and the dimension of MDS were determined empirically to maximize the correlation between subjective ratings and psychophysical measures. The concept for determination of the MDS-based psychophysical similarity measure is illustrated in Fig. 2.

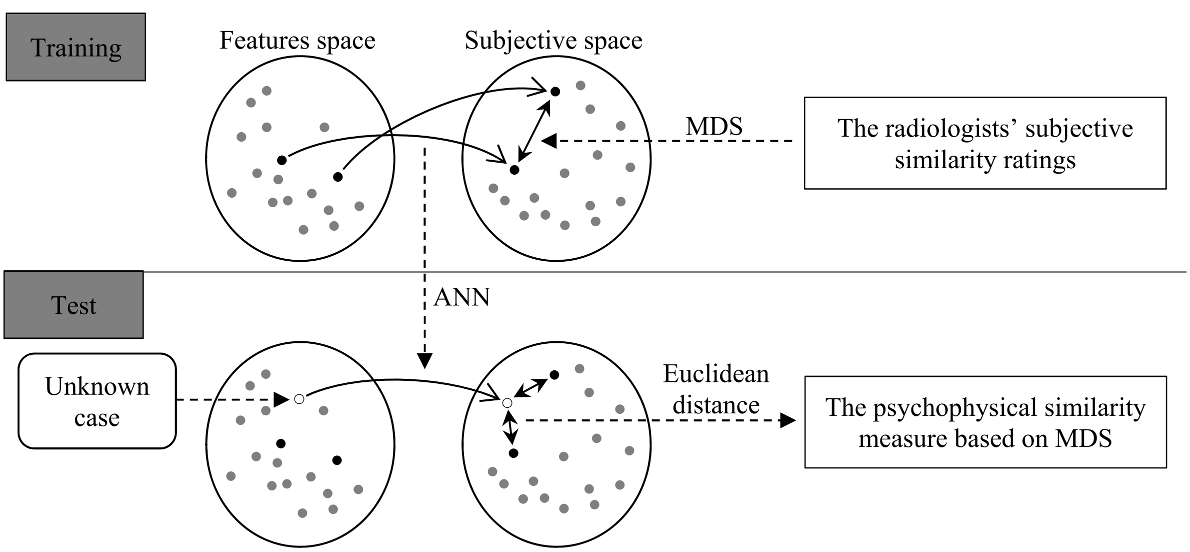

Figure 2. Conceptual diagram of determination of the psychophysical similarity measure based on MDS

3. RESULTS

The conventional psychophysical similarity measure was moderately correlated with the subjective similarity ratings (correlation coefficient: 0.68 [95% confidence interval: 0.64-0.72]), whereas the psychophysical measure based on MDS was highly correlated (correlation coefficient: 0.81 [95% confidence interval: 0.78-0.83]). The relationship between the subjective similarity ratings and the conventional measure is shown in Fig. 3, whereas that between the subjective ratings and the MDS-based measure is shown in Fig. 4.

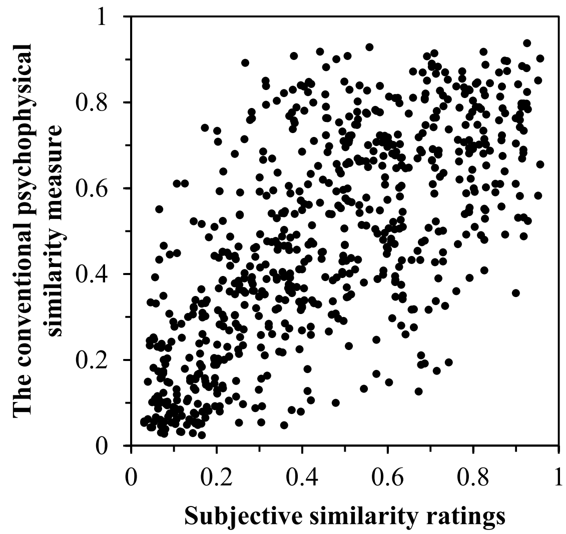

Figure 3. Relationship between subjective similarity ratings and the conventional psychophysical similarity measure

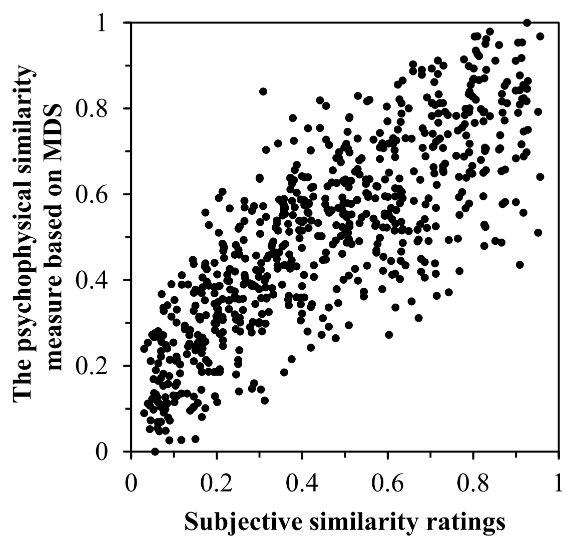

Figure 4. Relationship between subjective similarity ratings and the psychophysical similarity measure based on MDS

When *n* most similar images are selected, *precision* and *recall* are defined as follow.

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (2)$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (3)$$

The selection of a mass that has the same pathologic type as the index mass was considered the true positive selection. Therefore, the *precision* represents the fraction of selected masses that have the same pathologic type, while the *recall* represents the fraction of all the masses with the same pathologic type in the database that were selected. In order to compare with the result by using the subjective rating, similar images were selected from the 26 images used above (excluding a test case); therefore, there are at most 2 images out of 26 images that have the same pathologic type as the index image. Fig. 5 shows the average precision-recall curves by a leave-one-out test method when *n* was varied from1 to 26. The dotted curve shows the result by using the average subjective ratings. Similarly, Fig. 6 shows the results when only the concordance of benignity or malignancy was considered. In both evaluations, the MDS-based psychophysical measure obtained the higher performance than the conventional psychophysical similarity measure, although the results by using subjective ratings were superior.

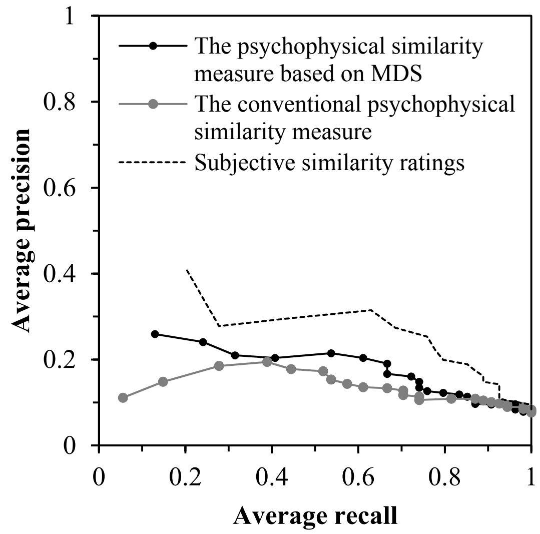

Figure 5. Precision-Recall curve of pathologic types

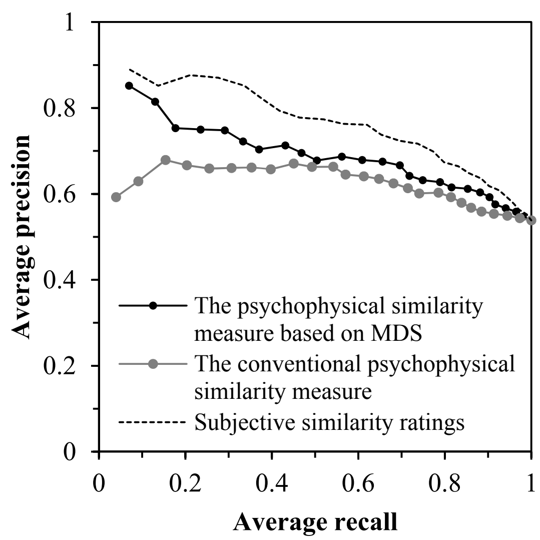

Figure 6. Precision-Recall curve of benign and malignant

4. CONCLUSIONS

In this study, we proposed a new psychophysical similarity measures based on MDS for retrieval of similar images of breast masses on mammograms, and examined its usefulness. The psychophysical similarity measure based on MDS was highly correlated with the radiologists' subjective similarity ratings, and the result indicates that the psychophysical similarity measure based on MDS can be useful in the retrieval of similar images. In the selection of similar images, precision and recall were higher when MDS-based psychophysical measure was used than when conventional measure was used. These results also indicate that a potential usefulness of the proposed measure in retrieval of reliable reference images.

ACKNOWLEDGMENT

Authors are grateful to following physicians for their contribution in providing the subjective similarity data: T. Horiba MD, M. Kato, MD, N. Kobayashi, MD, T. Morita, MD, M. Nishikawa, MD, T. Niwa, MD, M. Sassa, MD. This study was partly supported by JSPS Grant-in-Aid for Young Scientists (B) (No.21791179), and MEXT Grant-in-Aid for Scientific Research on Innovative Areas (No.21103004).

REFERENCES

[1] Matsuda, T., Marugame, T., Kamo, K. I., Katanoda, K., Ajiki, W., Sobue, T., The Japan Cancer Surveillance Research Group, "Cancer Incidence and Incidence Rates in Japan in 2006: Based on Data from 15 Population-based Cancer Registries in the Monitoring of Cancer Incidence in Japan (MCIJ) Project," Japanese Journal of Clinical Oncology, 42, 139-147 (2012).

[2] Chan, H. P., Sahiner, B., Helvie, M. A., Petrick, N., Roubidoux, M. A., Wilson, T. E., Adler, D. D., Paramagul, C., Newman, J. S., Sanjay-Gopal, S., "Improvement of Radiologists' Characterization of Mammographic Masses by Using Computer-aided Diagnosis: An ROC Study," Radiology, 212(3), 817-827 (1999).

[3] Muramatsu, C., Li, Q., Suzuki, K., Schmidt, R. A., Shiraishi, J., Newstead, G. M., Doi, K., "Investigation of psychophysical measure for evaluation of similar images for mammographic masses: Preliminary results," Medical Physics, 32(7), 2295-2304 (2005).

[4] Muramatsu, C., Nishimura, K., Oiwa, M., Shiraiwa, M., Endo, T., Doi, K., Fujita, H., "Correspondence among subjective and objective similarities and pathologic types of breast masses on digital mammography," Proc. IWDM 2012, 7361, 450-457 (2012).

[5] Kruskal, J. B., "Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis," Psychometrika, 29(1), 1-27 (1964).