

Evaluation of Objective Similarity Measures for Selecting Similar Images of Mammographic Lesions

Ryohei Nakayama,¹ Hiroyuki Abe,¹ Junji Shiraishi,¹ and Kunio Doi¹

The purpose of this study was to investigate four objective similarity measures as an image retrieval tool for selecting lesions similar to unknown lesions on mammograms. Measures A and B were based on the Euclidean distance in feature space and the psychophysical similarity measure, respectively. Measure C was the sequential combination of B and A, whereas measure D was the sequential combination of A and B. In this study, we selected 100 lesions each for masses and clustered microcalcifications randomly from our database, and we selected five pairs of lesions from 4,950 pairs based on all combinations of the 100 lesions by use of each measure. In two observer studies for 20 mass pairs and 20 calcification pairs, six radiologists compared all combinations of 20 pairs by using a two-alternative forced-choice method to determine the subjective similarity ranking score which was obtained from the frequency with which a pair was considered as more similar than the other 19 pairs. In both mass and calcification pairs, pairs selected by use of measure D had the highest mean value of the average subjective similarity ranking scores. The difference between measures D and A ($P = 0.008$ and 0.024), as well as that between measures D and B ($P = 0.018$ and 0.028) were statistically significant for masses and microcalcifications, respectively. The sequential combination of the objective similarity measure based on the Euclidean distance and the psychophysical similarity measure would be useful in the selection of images similar to those of unknown lesions.

KEY WORDS: Similarity measure, similar image, mass, clustered microcalcifications, mammogram

INTRODUCTION

In the interpretation of medical images, radiologists attempt to make diagnostic decisions based on the medical knowledge derived from viewing many clinical images over the years through education, training, and clinical practice. It is commonly known that, when a radiologist

may encounter a new, unknown case in daily clinical work, he/she may occasionally search for clinical images with known pathology similar to that of the unknown case by reviewing images in previous clinical cases, teaching files, and textbooks. Therefore, the presentation of similar images would be useful and would have the potential to improve radiologists' performance in the differential diagnosis of lesions in clinical images.¹⁻⁴

In order to develop a useful tool for selecting similar images to be used as a diagnostic aid, many investigators have studied content-based or feature-based image retrieval methods.⁵⁻¹⁷ However, these retrieval methods did not take into account radiologists' subjective impression of similarity when two images are compared. If retrieved images were not really similar to an unknown lesion visually for clinical purposes, they would not be useful for radiologists in the differential diagnosis of the unknown lesion. Therefore, Li et al.¹⁸ and Muramatsu et al.¹⁹⁻²³ have studied a psychophysical similarity measure, as an image retrieval tool, which was determined by use of an artificial neural network (ANN) for learning the

¹From the Kurt Rossmann Laboratories for Radiologic Image Research, Department of Radiology, The University of Chicago, 5841 South Maryland Avenue, MC 2026, Chicago, IL, 60637, USA.

Correspondence to: Ryohei Nakayama, Kurt Rossmann Laboratories for Radiologic Image Research, Department of Radiology, The University of Chicago, 5841 South Maryland Avenue, MC 2026, Chicago, IL, 60637, USA; tel: +1-773-8345093; fax: +1-773-7020371; e-mail: nakayama@clin.medic.mie-u.ac.jp

Copyright © 2010 by Society for Imaging Informatics in Medicine

Online publication 30 March 2010

doi: 10.1007/s10278-010-9288-3

relationship between radiologists' subjective similarity ratings and the objective features of lesions. They showed that the correlation coefficients ($r=0.72, 0.74,$ and 0.71 for nodules on low-dose CT and masses and clustered microcalcifications on mammograms, respectively) between radiologists' subjective similarity ratings and psychophysical similarity measures were greater than those ($r=0.60, 0.60,$ and 0.58 for nodules on low-dose CT and masses and clustered microcalcifications on mammograms, respectively) between radiologists' subjective similarity ratings and objective similarity measures based on the Euclidean distance in feature space that was frequently used in many studies. Their results indicated that similar images selected based on the psychophysical similarity measures would be more similar in terms of radiologists' visual perception than those selected based on feature space. However, it appears that the psychophysical similarity measures were not highly accurate as a reliable objective similarity measure for selecting similar images because the correlation coefficients were less than 0.80, i.e., they were not extremely high.

In this study, therefore, we investigated new objective similarity measures based on both the Euclidean distance in feature space and the psychophysical similarity measure. In order to evaluate the usefulness of these measures, we selected pairs of masses and pairs of clustered microcalcifications on mammograms by using four different measures. We conducted two observer studies based on a two-alternative forced-choice (2AFC) method²⁴ for mass pairs and for calcification pairs, for comparison of subjective similarities in terms of radiologists' visual perception on pairs of images selected by use of different measures.

MATERIALS AND METHODS

The use of the following database and the participation of radiologists in the observer study were approved by the Institutional Review Board at our medical center. Informed consent for this observer study was obtained from all observers.

In this study, we investigated four objective similarity measures (A, B, C, and D) based on the Euclidean distance in feature space and the psychophysical similarity measure determined by

the ANN. In our previous studies,^{19,22,23} 50 images including 25 benign and 25 malignant lesions were first selected as representative lesions for both mass and calcification studies by an attending breast radiologist to include various sizes and types of lesions. Three hundred pairs were created by the combination of each representative lesion and six images (three benign and three malignant lesions) selected subjectively by consensus of three investigators to include pairs with a wide range of similarities. Ten breast radiologists provided their subjective similarity ratings for the 300 mass pairs and the 300 calcification pairs. For specific image features considered in both the Euclidean distance and the ANN, we employed the combination of six and seven objective features for masses and clustered microcalcifications, respectively, which provided the highest correlation coefficients between the average subjective similarity ratings and psychophysical similarity measures.^{19,22,23} The six features for masses included the degree of irregularity, the full width at half maximum of a cumulative modified radial gradient histogram, the radial gradient index, the minor-to-major axis ratio of an ellipse fitted to the outline of the mass, the edge contrast, and the standard deviation of pixel values.^{19,23} On the other hand, the seven features for clustered microcalcifications included the circularity of the cluster, the number of microcalcifications per unit area, the mean effective diameter of microcalcifications, the standard deviation of the effective diameters of microcalcifications, the mean contrast of microcalcifications, the standard deviation of contrasts of microcalcifications, and the standard deviation of the shape irregularities of microcalcifications.^{22,23} Measures A and B were based on the Euclidean distance in feature space and the psychophysical similarity measure, respectively. Measure C was the sequential combination of B and A, which was derived first based on the psychophysical similarity measure and then the Euclidean distance in feature space, whereas measure D was the sequential combination of A and B, which was derived based on the Euclidean distance in feature space and then the psychophysical similarity measure.

Databases

To compare the usefulness of four different measures as an image retrieval tool, we used pairs

of masses and pairs of clustered microcalcifications on mammograms which were obtained from the Digital Database for Screening Mammography developed by the University of South Florida.²⁵ Our database for masses consisted of 1,568 regions of interest (ROIs), including 840 benign and 728 malignant masses.²³ The size of the ROI was 5×5 cm (pixel size 100 μm), centered at each mass. On the other hand, our database for clustered microcalcifications consisted of 1,101 ROIs, including 644 benign and 457 malignant clustered microcalcifications.²³ The size of the ROI was 3×3 cm (pixel size 50 μm), centered at each clustered microcalcification. All lesions were proved by biopsy. The contrast and the density level in each ROI were manually adjusted to an appropriate level by an attending breast radiologist.

Selection of Pairs of Images

The pairs of images for masses and those for clustered microcalcifications were selected for each of the observer studies by use of the method described below. We first removed 300 ROIs used for training the ANN,^{19,22,23} which was then applied to the determination of psychophysical similarity measures for all of pairs of images used in this study. One hundred ROIs were selected randomly from the remaining ROIs (1,268 and 801 for masses and clustered microcalcifications, respectively) such that only one ROI would be selected from the same patient. For the selected 100 ROIs, 4,950 pairs were created by all possible combinations of two different ROIs. Pairs of ROIs with the highest similarity measures were then selected for an observer study by use of four different measures. For measure A, five pairs with the five highest similarity measures based on the Euclidean distances in feature space were selected from the 4,950 pairs. For measure B, five pairs with the five highest psychophysical similarity measures were selected from the 4,950 pairs. For measure C, a pair with the highest psychophysical similarity measure was preselected in 99 pairs created by the combinations of one ROI and the other 99 ROIs. This procedure was repeated for all of the selected 100 ROIs. Subsequently, five pairs with the five highest similarity measures based on the Euclidean distances were selected from the preselected 100 pairs. For measure D, a pair with the highest similarity measure based on the

Euclidean distance was preselected in 99 pairs created by the combinations of one ROI and the other 99 ROIs. This procedure was repeated for all of the selected 100 ROIs. Subsequently, five pairs with the five highest psychophysical similarity measures were selected from the preselected 100 pairs. Here, five pairs for each measure were selected such that the same ROI would not be selected again as another ROI in different pairs obtained with the same measure.

Observer Study

We conducted two observer studies for 20 mass pairs and for 20 calcification pairs, for comparison of subjective similarities in terms of radiologists' visual perception on pairs of ROIs selected by use of the four different measures. The 2AFC method, known as a paired comparison method, was employed in the observer study because it is a sensitive method for the distinction of a small difference in the comparison of two similar patterns.²⁴ In the observer study, two pairs of lesions were displayed on a high-resolution liquid crystal display monitor (ME511L/P4, 21.3 in., $2,048 \times 2,560$ pixels, 410 cd/m^2 luminance; Totoku Electric Co., Ltd.) with one pair above and another pair below, as shown in Figure 1. The observer was asked to compare the similarity of the two pairs and to select the pair considered more similar than the other pair. During the observer study, each pair was compared to all of the other 19 pairs one by one. The frequency with which a pair was selected as the more similar pair was considered as the subjective similarity ranking score for the pair; the maximum and the minimum score would be 19 and 0, respectively. The subjective similarity ranking scores indicate the relative rankings of similarities among the 20 pairs selected by four different measures.

Six observers, including three attending breast radiologists and three breast-imaging fellows, participated independently in the observer study. The instructions to the observers were the following: (1) the purpose of this study is to obtain experimental data for subjective impression of similarity for pairs of masses (and pairs of clustered microcalcifications in the second session) on mammograms selected by four computerized methods. (2) Two pairs of images are displayed on a monitor. You are asked to compare the similarity

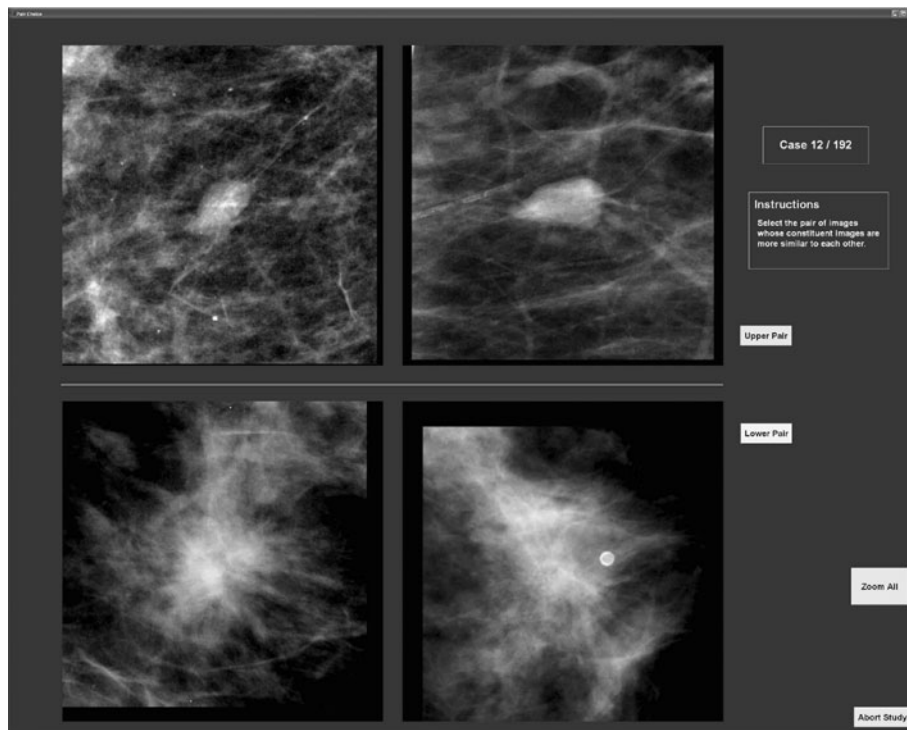


Fig 1. Observer interface for obtaining subjective similarity ranking scores based on the 2AFC method.

of one pair above with that of another pair below, regarding the overall impression for diagnosis. Click on the one pair that is more similar than the other. (3) A training session including two comparisons of pairs of lesions is provided at the beginning of the study. (4) There is no time limit.

RESULTS

Figure 2 shows the relationship between the objective similarity measures based on the Euclidean distance and the psychophysical similarity measures for 20 mass pairs selected by the four different measures. The mass pairs selected by use of measure A tended to have high objective similarity measures based on the Euclidean distances and relatively low psychophysical similarity measures, whereas those by measure B tended to have relatively low objective similarity measures based on the Euclidean distances, but high psychophysical similarity measures. The pairs selected by use of measures C and D were distributed between the pairs for measures A and B. The pairs for measure C were distributed near the pairs for measure A, whereas the pairs for measure D were

distributed near the pairs for measure B. It should be noted that there is a noticeable difference among four groups of pairs of masses selected as “most similar” based on the four different methods. Figure 3a, b shows the relationships for the

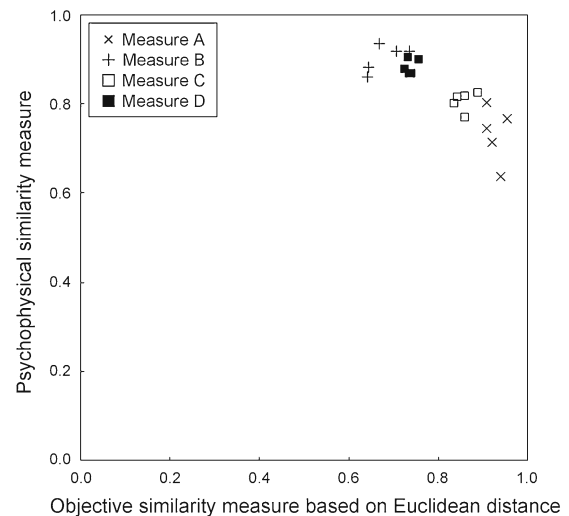


Fig 2. Relationship between objective similarity measure based on the Euclidean distance in feature space and psychophysical similarity measure for 20 mass pairs selected by four different measures.

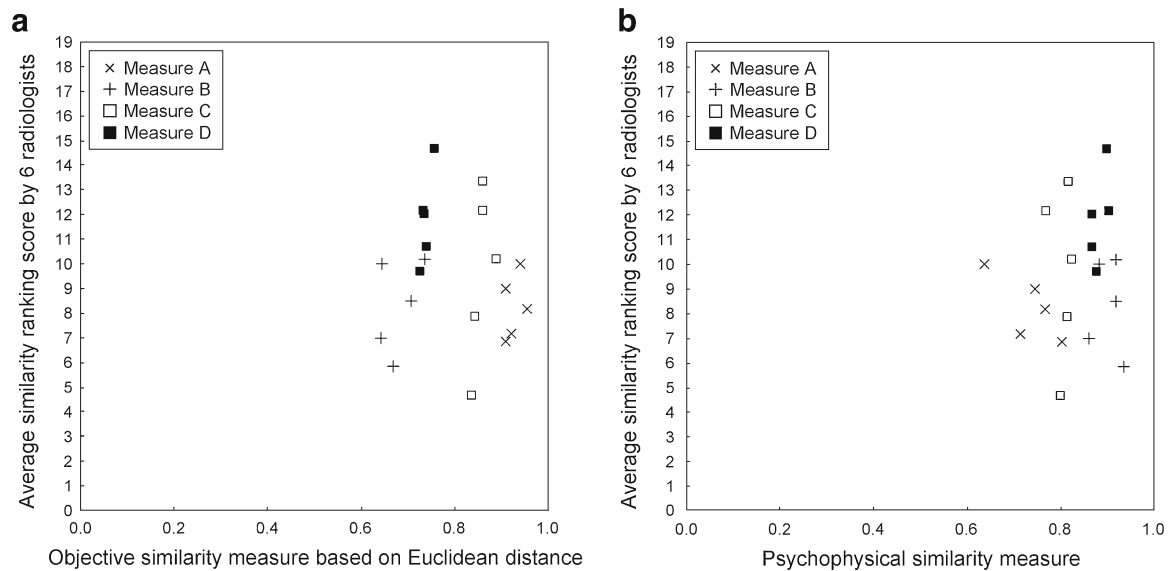


Fig 3. a Relationship for average similarity ranking score of each mass pair by six radiologists with objective similarity measure based on the Euclidean distance. b Relationship for average similarity ranking score with psychophysical similarity measure.

average subjective similarity ranking score of mass pairs by six radiologists with the objective similarity measure based on the Euclidean distance and also with the psychophysical similarity measure, respectively. Table 1 shows the mean values and the standard deviations of the average subjective similarity ranking scores for four groups of mass pairs selected by use of different measures. Although there was a large variation in the average similarity ranking scores for each measure, the mean value of the average similarity ranking scores for measure D was greater than those for the three other measures. On the other hand, the mean value of the average similarity ranking scores for measure A was lower than those for the other measures. These results indicated that the mass pairs selected by measure D were more similar, on average, in terms of radiologists' visual perception, than those by the other measures. Table 2 shows P values for the difference in the

average similarity ranking scores obtained by use of two different measures. A statistical analysis was performed with use of Student's t test based on the average similarity ranking score for each pair obtained by six radiologists. The difference ($P=0.008$) between measures D and A and that ($P=0.018$) between measures D and B were statistically significant. Figures 4 and 5 show the 20 mass pairs obtained by use of the four different measures, together with the average subjective similarity ranking score in bold (ranking on objective similarity measures based on the Euclidean distance in 4,950 pairs/objective similarity measure and also ranking on psychophysical similarity measures in 4950 pairs/psychophysical similarity measure) for each pair. The first pair for measure D in Figure 5 had the highest average similarity ranking score, whereas the fifth pair for measure C had the lowest average similarity ranking score.

Table 1. Mean Values and Standard Deviations of Average Subjective Similarity Ranking Scores of Mass Pairs by Six Radiologists for Each Measure

	Mean \pm SD
Measure A	8.23 \pm 1.30
Measure B	8.30 \pm 1.88
Measure C	9.63 \pm 3.47
Measure D	11.83 \pm 1.89

Table 2. P values for the Difference in the Average Subjective Similarity Ranking Scores of Mass Pairs Selected by Two Different Measures

	Measure A	Measure B	Measure C
Measure B	<0.950		
Measure C	<0.423	<0.472	
Measure D	<0.008	<0.018	<0.248

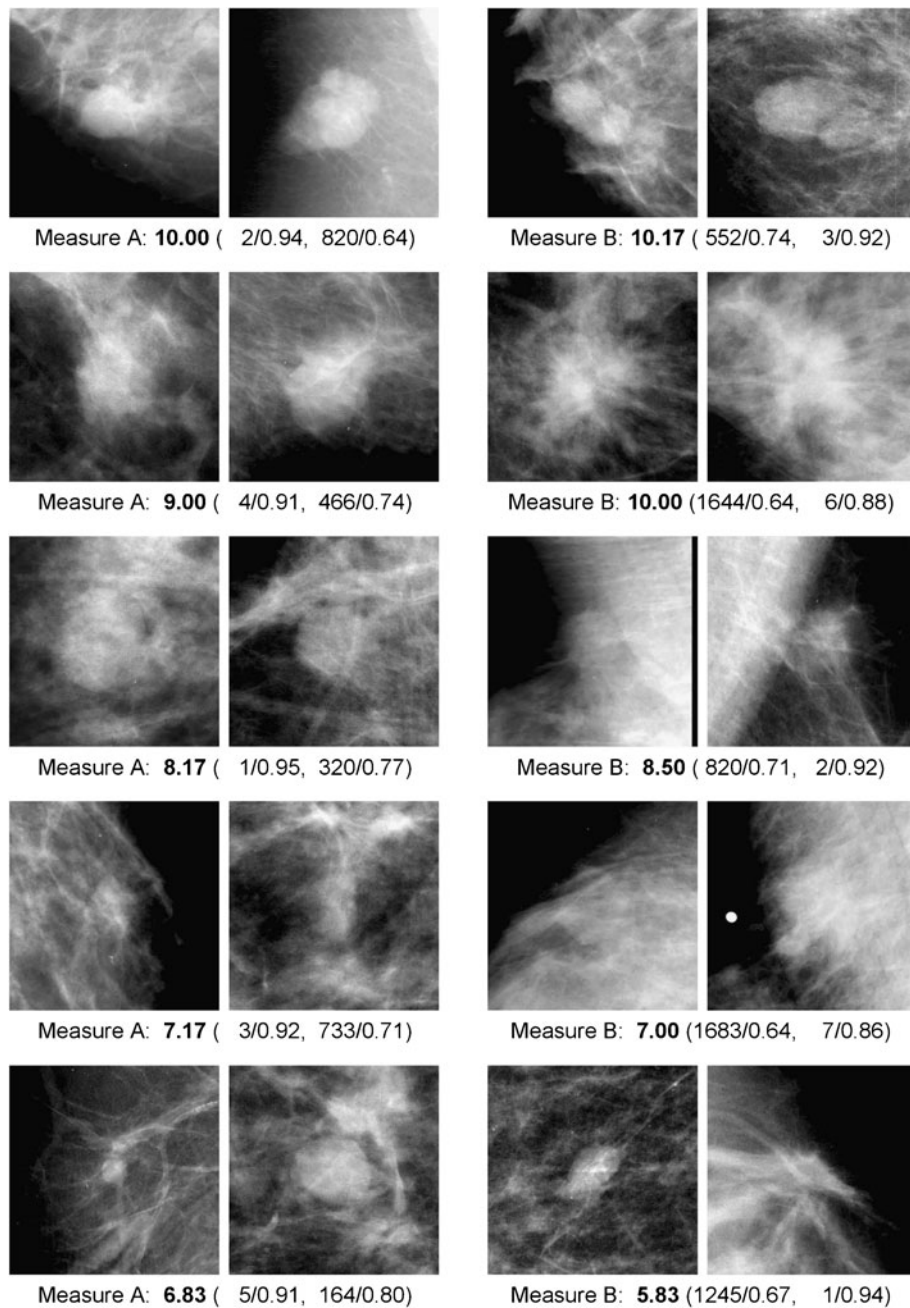


Fig 4. Mass pairs for measures A and B and the average subjective similarity ranking score in *bold* (ranking on objective similarity measures based on the Euclidean distance in 4,950 pairs/objective similarity measure, ranking on psychophysical similarity measures in 4,950 pairs/psychophysical similarity measure) for each pair.

Figure 6 shows the relationship between the objective similarity measure based on the Euclidean distance and the psychophysical similarity measure for 20 calcification pairs selected by the four measures. Although there was a small overlap in the distributions of calcification pairs among the

four measures, the calcification pairs for each of the measures tended to be distributed in a way similar to those for the mass pairs in Figure 2. Figure 7a, b shows the relationships for the average subjective similarity ranking score of calcification pairs to the objective similarity

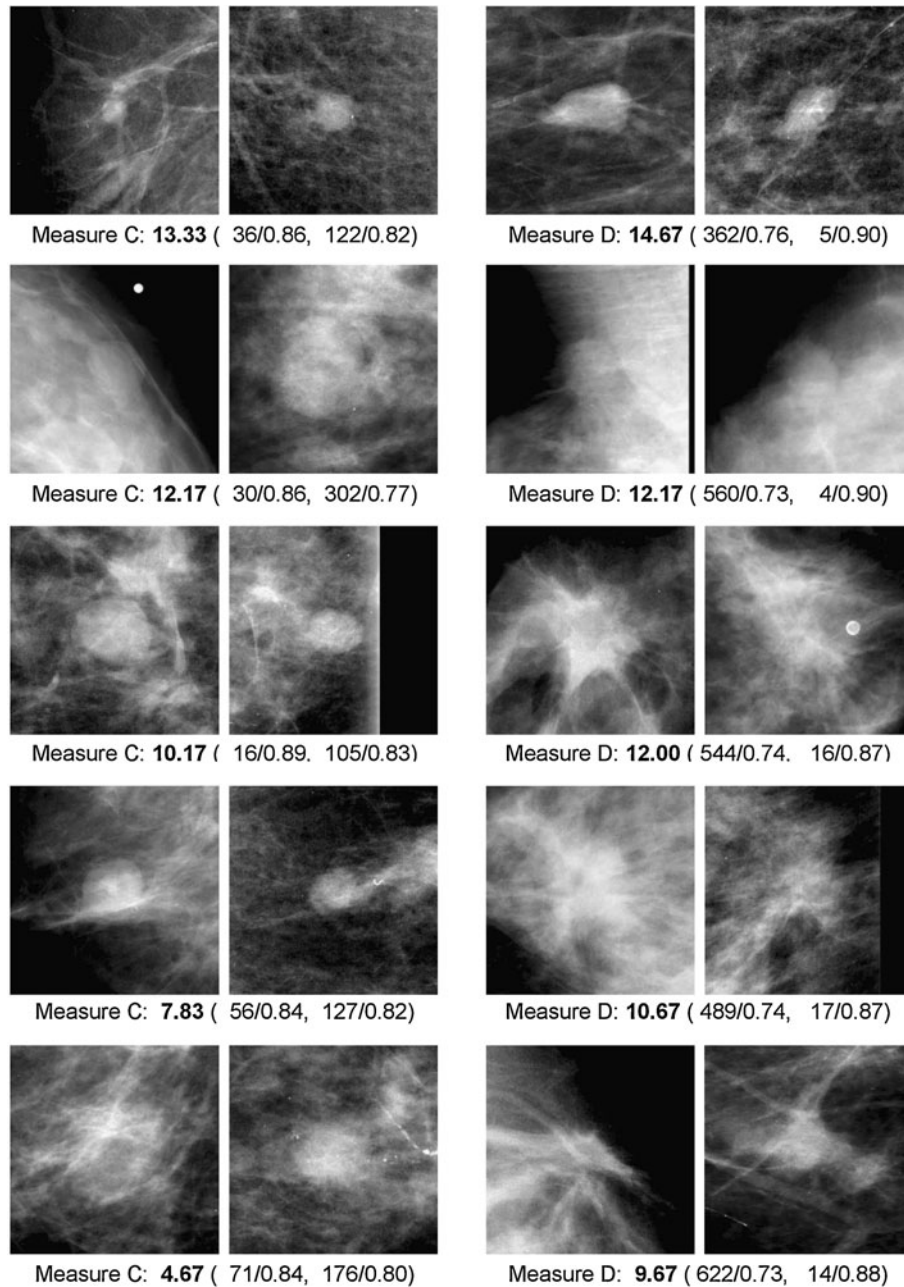


Fig 5. Mass pairs for measures C and D and the average subjective similarity ranking score in *bold* (ranking on objective similarity measures based on the Euclidean distance in 4,950 pairs/objective similarity measure, ranking on psychophysical similarity measures in 4,950 pairs/psychophysical similarity measure) for each pair.

measure based on the Euclidean distance and to the psychophysical similarity measure, respectively. Table 3 shows the mean values and the standard deviations of the average subjective similarity ranking scores of calcification pairs for each measure. The calcifications pairs for measure D had the highest average subjective similarity

ranking scores, whereas those for measure A had the lowest average similarity ranking scores; these results were the same as those for masses. Table 4 shows P values for the difference in the average similarity ranking scores obtained by use of two different measures. The difference ($P=0.024$) between measures D and A and that difference

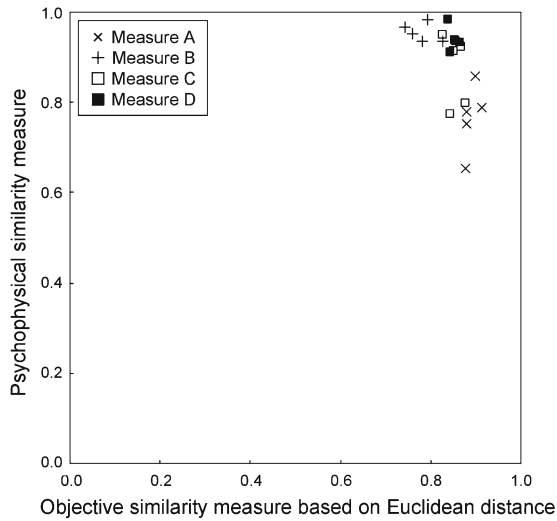


Fig 6. Relationship between objective similarity measure based on the Euclidean distance in feature space and psychophysical similarity measure for 20 calcification pairs selected by four different measures.

($P=0.028$) between measures D and B were statistically significant. Figure 8 shows calcification pairs with the highest average subjective similarity ranking score in each objective similarity measure, together with the average subjective similarity ranking score in bold (ranking on objective similarity measures based on the Euclidean distance in 4,950 pairs/objective similarity measure and also ranking on psychophysical

similarity measures in 4,950 pairs/psychophysical similarity measure) for each pair. The pairs with very high objective similarity measures both for the Euclidean distance and the ANN tended to have high average subjective similarity ranking scores in measures C and D.

DISCUSSION

In both observer studies for mass pairs and calcification pairs, the mean values of the average subjective similarity ranking scores for measure B were greater than those for measure A, although the difference between measures A and B was not statistically significant in this study. This result tended to be consistent with the results presented by Li et al.¹⁸ and Muramatsu et al.,^{19,22,23} where the correlation coefficient of radiologists' subjective similarity ratings with psychophysical similarity measures was greater than that with objective similarity measures based on the Euclidean distance. These results may indicate that the psychophysical similarity measure is a better tool in retrieving similar images than is the objective similarity measure based on the Euclidean distance.

The mean values of the average similarity ranking scores for measures C and D were greater than those for measures A and B. The mean value of the average similarity ranking scores for

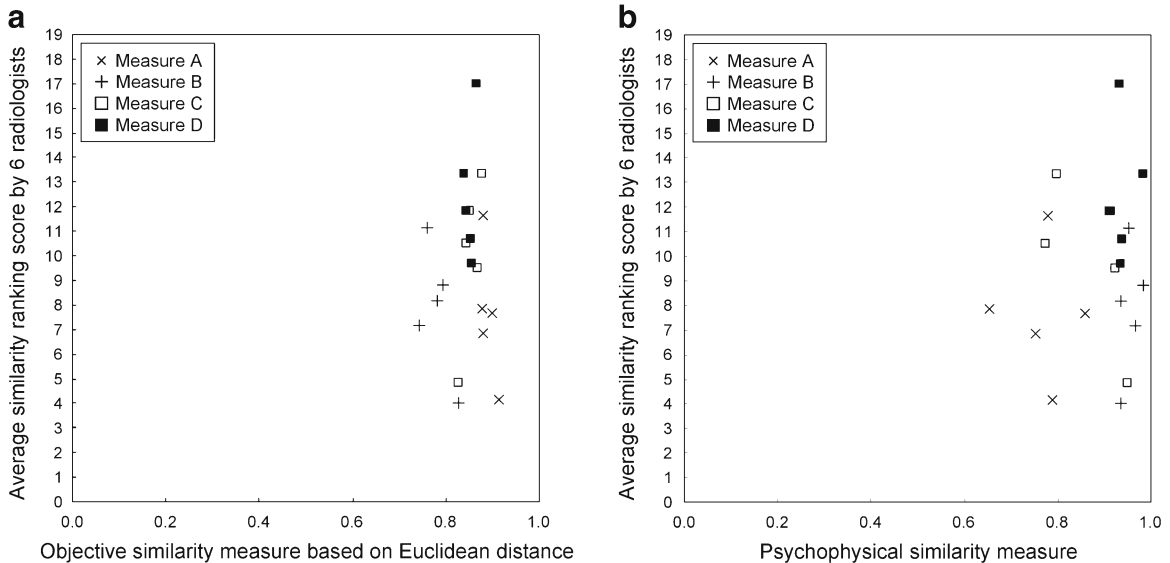


Fig 7. a Relationship for average similarity ranking score of each calcification pair by six radiologists with objective similarity measure based on the Euclidean distance. b Relationship for average similarity ranking score with psychophysical similarity measure.

Table 3. Mean Values and Standard Deviations of Average Subjective Similarity Ranking Scores of Calcification Pairs by Six Radiologists for Each Measure

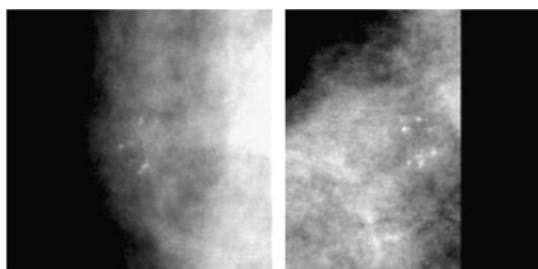
	Mean \pm SD
Measure A	7.63 \pm 2.69
Measure B	7.87 \pm 2.62
Measure C	10.00 \pm 3.23
Measure D	12.50 \pm 2.86

measure D was greater than that for C. For measure D, the pairs with comparable physical characteristics were first preselected by use of an objective similarity measure based on the Euclidean distance, and thus, the subsequent selection of pairs with high psychophysical similarity measures would be more reliable because inadequate pairs which may not be similar due to a large difference in physical characteristics were removed initially. Therefore, we believe that the pairs selected by measure D would be more similar in terms of radiologists' visual perception than those by measure B because measure B was improved substantially by the sequential combination with measure A. With measure C, on the other hand, the pairs were first preselected by use of a psychophysical similarity measure, and thus, some pairs with high objective similarity measures, which would be located closely in feature space, would have been removed, and the subsequent selection of pairs may provide pairs with different physical characteristics. Therefore, we believe that the pairs for measure D would be more similar subjectively than those for measure C.

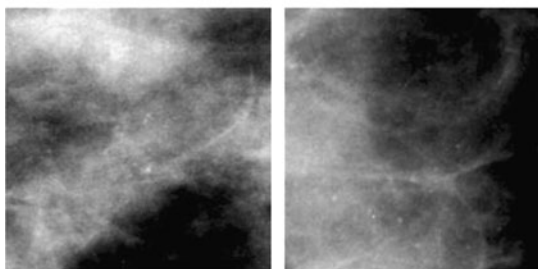
The implementation of selecting similar images by use of measure D in clinical situations can be illustrated in the example described below. When a radiologist encounters a new, unknown case in daily clinical practice at a breast clinic, a search engine would determine first the objective similarity measures based on the Euclidean distance in feature space for all of the combinations for the

Table 4. P Values for the Difference in the Average Subjective Similarity Ranking Scores of Calcification Pairs Selected by Two Different Measures

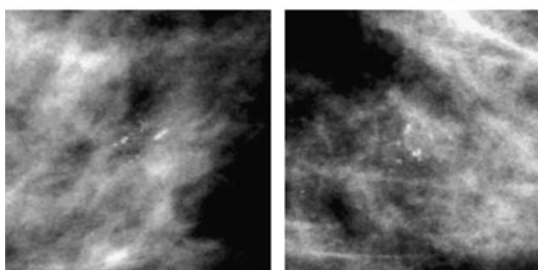
	Measure A	Measure B	Measure C
Measure B	<0.893		
Measure C	<0.244	<0.285	
Measure D	<0.024	<0.028	<0.231



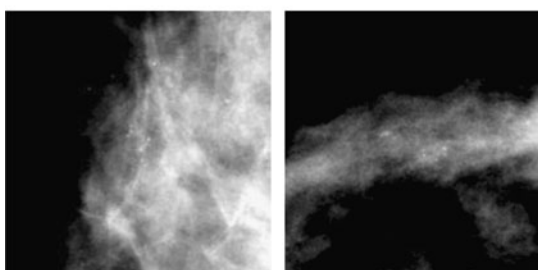
Method A: **11.67** (5/0.88, 447/0.78)



Method B: **11.17** (445/0.76, 25/0.95)



Method C: **13.33** (7/0.88, 370/0.80)



Method D: **17.00** (13/0.87, 48/0.93)

Fig 8. Calcification pairs with the highest average subjective similarity ranking score in each objective similarity measure and the average subjective similarity ranking score in *bold* (ranking on objective similarity measures based on the Euclidean distance in 4,950 pairs/objective similarity measure, ranking on psychophysical similarity measures in 4,950 pairs/psychophysical similarity measure) for each pair.

unknown case with all of the known benign/malignant cases in the database available in the clinic, which may include a large number of cases such as 1,000 benign cases and 1,000 malignant cases stored in a picture archiving and communi-

cation system. The search engine would then select a certain pre-selected number of cases such as the top 100 pairs, each for benign and malignant cases, with higher objective similarity measures; these pairs would be subjected to determination of the psychophysical similarity measures by use of the trained ANN. Finally, the radiologist may indicate a desired number of similar cases to be presented as an aid to his/her diagnosis, such as five cases each for benign/malignant cases. The search engine then could retrieve those cases with the five highest psychophysical similarity measures in each category to be presented as similar cases. It is likely that the cases selected would look more similar to the unknown case in question for radiologists in making their diagnostic decision than other cases which might be selected by the three other measures, A, B, or C.

There are some limitations in this study. One limitation is that the number of pairs for each objective similarity measure was small in the observer study because the time required for a radiologist has to be limited to an hour in one session. Another limitation is that four of six breast radiologists who participated in the observer study provided their subjective similarity ratings for the 300 mass pairs and the 300 calcification pairs in our previous studies. However, we believe that the bias due to this overlap would be minimal because for training the ANN, the average subjective similarity ratings were obtained by ten breast radiologists.

CONCLUSION

In both mass and calcification pairs, pairs selected by use of measure D, which was the sequential combination of the objective similarity measure based on the Euclidean distance in feature space with the psychophysical similarity measure, had the highest mean value of the average subjective similarity ranking scores. Measure D would be useful in the selection of images similar to those of unknown masses or clustered microcalcifications on mammograms.

ACKNOWLEDGMENTS

We are grateful to Kirti Kulkarni, MD, Charlene Sennett, MD, Rajshri Shah, MD, Steven G. Thiel, MD, and Akiko Shimauchi, MD, for participating in the observer study; to Chisako Muramatsu, PhD, for her useful suggestions; and to

Elisabeth Lanzl for improving the manuscript. K. Doi is a shareholder in Hologic/R2 Technology, Inc., Los Altos, CA, USA. CAD technologies developed in the Kurt Rossmann Laboratories have been licensed to R2 Technology, Deus Technologies, Riverain Medical Group, Mitsubishi Space Software Co., Median Technologies, General Electric Corporation, and Toshiba Corporation. It is the policy of The University of Chicago that investigators disclose publicly actual or potential significant financial interests that may appear to affect research activities or that may benefit from research activities.

REFERENCES

1. Doi K: Current status and future potential of computer-aided diagnosis in medical imaging. *Br J Radiol* 78:s3-s19, 2005
2. Doi K: Computer-aided diagnosis in medical imaging: historical review, current status and future potential. *Comput Med Imaging Graph* 31:198-211, 2007
3. Doi K: Computer-aided diagnosis moves from breast to other systems. *Diagn Imag* 29:37-40, 2007
4. Kumazawa S, Muramatsu C, Li Q, et al: An investigation of radiologists' perception of lesion similarity: observations with paired breast masses on mammograms and paired lung nodules on CT images. *Acad Radiol* 15:887-894, 2008
5. Bucci G, Cagnoni S, De Dominicis R: Integrating content-based retrieval in a medical image reference database. *Comput Med Imaging Graph* 20:231-241, 1996
6. Wong ST, Huang HK: Design methods and architectural issues of integrated medical image data base systems. *Comput Med Imaging Graph* 20:285-299, 1996
7. Swett HA, Fisher PR, Cohn AI, et al: Expert system controlled image display. *Radiology* 172:487-493, 1989
8. Swett HA, Mutalik PG, Neklesa VP, et al: Voice-activated retrieval of mammography reference images. *J Digit Imaging* 11:65-73, 1998
9. Qi H, Snyder WE: Content-based image retrieval in picture archiving and communications systems. *J Digit Imaging* 12(2 Suppl 1):81-83, 1999
10. Sklansky J, Tao EY, Bazargan M, et al: Computer-aided, case-based diagnosis of mammographic regions of interest containing microcalcifications. *Acad Radiol* 7:395-405, 2000
11. Giger ML, Huo Z, Vyborny CJ, et al: Intelligent CAD workstation for breast imaging using similarity to known lesions and multiple visual prompt aids. *Proc SPIE* 4684:768-773, 2002
12. Sinha U, Kangaroo H: Principal component analysis for content-based image retrieval. *Radiographics* 22:1271-1289, 2002
13. Aisen AM, Broderick LS, Winer-Muram H, et al: Automated storage and retrieval of thin-section CT images to assist diagnosis: system description and preliminary assessment. *Radiology* 228:265-270, 2003
14. Kawata Y, Niki N, Ohmatsu H, et al: Example-based assisting approach for pulmonary nodule classification in three-dimensional thoracic computed tomography images. *Acad Radiol* 10:1402-1415, 2003
15. El-Naqa I, Yang Y, Galatsanos NP, et al: A similarity learning approach to content-based image retrieval: application

to digital mammography. *IEEE Trans Med Imaging* 23:1233–1244, 2004

16. Zheng B, Lu A, Hardesty LA, et al: A method to improve visual similarity of breast masses for an interactive computer-aided diagnosis environment. *Med Phys* 33:111–117, 2006

17. Nakayama R, Watanabe R, Namba K, et al: An improved computer-aided diagnosis scheme using the nearest neighbor criterion for determining histological classification of clustered microcalcifications. *Methods Inf Med* 46:716–722, 2007

18. Li Q, Li F, Shiraishi J, et al: Investigation of new psychophysical measures for evaluation of similar images on thoracic computed tomography for distinction between benign and malignant nodules. *Med Phys* 30:2584–2593, 2003

19. Muramatsu C, Li Q, Suzuki K, et al: Investigation of psychophysical measure for evaluation of similar images for mammographic masses: preliminary results. *Med Phys* 32:2295–2304, 2005

20. Muramatsu C, Li Q, Schmidt RA, et al: Experimental determination of subjective similarity for pairs of clustered

microcalcifications on mammograms: observer study results. *Med Phys* 33:3460–3468, 2006

21. Muramatsu C, Li Q, Schmidt RA, et al: Determination of subjective similarity for pairs of masses and pairs of clustered microcalcifications on mammograms: comparison of similarity ranking scores and absolute similarity ratings. *Med Phys* 34:2890–2895, 2007

22. Muramatsu C, Li Q, Schmidt RA, et al: Investigation of psychophysical similarity measures for selection of similar images in the diagnosis of clustered microcalcifications on mammograms. *Med Phys* 35:5695–5702, 2008

23. Muramatsu C: Investigation of similarity measures for selection of similar images in computer-aided diagnosis of breast lesions on mammograms. Ph.D. dissertation, The University of Chicago, Chicago, IL (ProQuest/UMI, Ann Arbor, MI, 2008)

24. Kendall M, Gibbons JD: Rank correlation methods, 5th edition. Oxford University Press, New York, 1990

25. Heath M, Bowyer K, Kopans D, et al: Current status of the digital database for screening mammography. In: *Digital mammography*. Dordrecht: Kluwer Academic, 1998