Machine Classification of Dental Images with Visual Search[1]

Dennis P. Carmody, PhD, Susan P. McGrath, PhD,[2] Stanley M. Dunn, PhD
Paul F. van der Stelt, DDS, PhD, Elly Schouten, MA

Rationale and Objectives. The authors performed this study to assess the performance of a computer-based classification system that uses gaze locations of observers to define the subspace for machine learning.

Materials and Methods. Thirty-two dental radiographs were classified by an expert viewer into four categories of disease of the periapical region: no disease (normal tooth), mild disease (widened periodontal ligament space), moderate disease (destruction of the lamina dura), and severe disease (resorption of bone in the periapical area). There were eight images in each category. Six observers independently viewed the images while their eye gaze position was recorded. They then classified the images into one of the four categories. A sample of image space was used as input to a machine learning routine to develop a machine classifier. Sample space was determined with three techniques: visual gaze, random selection, and constrained random selection. κ analyses were used to compare classification accuracies with the three sampling techniques.

Results. With use of the expert classification as a standard of reference, observers classified images with 57% accuracy, and the machine classified images with 84% accuracy by using the same gaze-selected features and image space. Results of κ analyses revealed mean values of 0.78 for gaze-selected sampling, 0.69 for random sampling, 0.68 for constrained random selection, and 0.44 for observers. The use of sample space selected with the visual gaze technique was superior to that selected with both random-selection techniques and by the observers.

Conclusion. Machine classification of dental images improves the accuracy of individual observers using gaze-selected image space.

Key Words. Diagnostic radiology, observer performance; images, interpretation; quality assurance.

Digital image processing of dental radiographs is a growing technology, and it is expected that diagnostic decision accuracy will be improved with the image-processing techniques available (1). Among the areas expected to improve are the diagnoses of caries, periodontitis, and periapical disease; the ability for three-dimensional viewing of the teeth and supporting structures; and the analyses of bone trabecular patterns for early detection of systemic disease (2). Several advantages of this technology include a lower radiation dose to the patient, an immediate availability of images without the need for chemical film processing, and the possibility of image enhancement and computer-aided feature extraction (3). Another possible advantage is the use of computer-based image classification systems with the digital images.

One of the main problems associated with computer-based image classification systems is the extraction of the

Acad Radiol 2001; 8:1239–1246

[1] From the Department of Psychology, Saint Peter's College, 2641 Kennedy Blvd, Jersey City, NJ 07306 (D.P.C.); the Department of Biomedical Engineering, Rutgers University, Piscataway, NJ (S.P.M., S.M.D.); the Department of Oral and Maxillofacial Radiology, Academic Center for Dentistry, Amsterdam, the Netherlands (P.F.v.d.S.); and the Faculty of Human Movement Sciences, Vrije Universiteit Amsterdam, the Netherlands (E.S.). Received April 10, 2001; revision requested April 25; revision received July 8; accepted August 13. S.P.M. supported in part by an Investigative Research Award from Naval Air Warfare Center. S.M.D. supported in part by National Institutes of Health grant AR08457. D.P.C. supported in part by a Faculty Fellowship from Saint Peter's College. **Address correspondence to** D.P.C.

[2] Current address: Thayer School of Engineering, Dartmouth College, Hanover, NH.

relevant information from the visual scene. Most current computer-based systems process the image on a pixel-by-pixel basis and then extract features to be used for differentiating among classes. Such an approach often results in the processing of a large amount of unnecessary information. If image space is reduced to a sample of highly informative areas and the less informative and redundant areas are eliminated, then the reduced sample space may provide sufficient accuracy for classifying cases as diseased or normal. Would a computer classification routine retain levels of accuracy if the sample space were reduced by means of a judicious selection of features?

One method used to select highly informative image details is to have an experienced observer preview the images while his or her line of sight is recorded. Then, only the areas of the images sampled by the observer would be processed with the computer program for classification. A brief review of findings of vision research in radiology will help identify the parameters of visual search that are relevant to develop the input for computer classification routines.

A review of research in the role of visual search in radiology will serve to determine the observer behaviors that are relevant to accuracy in the detection and classification of disease patterns. In the earliest works about visual search in chest radiology, investigators found that observers did not scan the entire chest radiograph before making a diagnostic decision (4–7). Further investigations of search patterns revealed that radiologic experience modified the scan paths of observers (8,9) and that the scan paths were related to the detection of disease targets (10). Later work showed the respective roles of central and peripheral vision in search (11) and the need for context in decision making (12). Some studies have shown that observers use comparison eye scanning between suspected target disease areas and reference background context (13–16). The results of another study (17) has demonstrated that the durations of fixations in search are related to the decision accuracy.

Results of visual search studies in mammography have shown some similarities to the scanning behavior found in chest radiology. One research group found that the relationship between long eye gaze dwell time and diagnostic accuracy varied with the object of search. False-positive diagnoses were associated with long dwell times when masses were the targets, and true-positive diagnoses were associated with long dwell times when microcalcifications were the targets (18). Other research groups concluded that gaze duration was a predictor of missed lesions at mammography, and observers with different degrees of experience were identified with gaze durations, scan paths, and detection times (19–21).

Results of visual search studies in conventional radiography have helped confirm that the measures of fixation locations and gaze durations were related to decision accuracy in the detection of localized abnormalities. Similar findings were reported for abdominal studies (22) and bone fractures (23,24). Studies of visual search in cross-sectional studies, including magnetic resonance imaging of the head (25) and computed tomography of the chest (26,27), revealed similar patterns of search behavior. On dental images, the effects of experience were shown as shifts in gaze locations from the visually conspicuous to the more cognitively conspicuous image areas (28). A general conclusion from the many studies of gaze location and duration is that local abnormalities are found when fixations are directed to image areas within a relatively short distance of the target and missed when fixations are located farther than 3–5 visual degrees from targets. In general, local abnormalities are missed when fixation durations are brief and detected when the durations are sufficiently long. Some abnormalities, however, are missed despite prolonged visual attention accompanied at times by revisitation of the gaze to the suspected areas.

Attempts have been made to alter the characteristics of the display to reduce the observer errors of false-negative findings. One empirical study in which two viewing conditions—sectioned viewing and full image viewing—were compared to determine the detection accuracy for lung nodules on chest images revealed that more nodules were reported when viewing isolated sections than when viewing full images; the frequency of false-positive findings, however, increased (13). One explanation for the findings was that the loss of referential background information, especially in the contralateral lungs, deprived observers of an appreciation of the normal context of the lungs; this was confirmed with follow-up studies (14). Another study (29) showed that localization of abnormalities within a region of interest while the entire image is left available for referencing led to improved detection performance.

Another intervention used to improve decision accuracy is computer-aided diagnosis (CAD). One research group developed a computerized method to detect the location of lung nodules on digital chest images, which led to a slight (although not statistically significant) improvement in the detection performance of radiologists who were alerted to potential target locations (30). The authors indicated that improvements in the computerized

scheme would increase the sensitivity and specificity of the program. In a follow-up study from the same research laboratory (31), digital chest images were interpreted by 77 radiologists, 27 radiology residents, and 42 nonradiologists first without and then with CAD. Detection accuracy improved for all observers, which led to the conclusion that CAD has the potential to improve the diagnostic accuracy in the detection of lung nodules on digital radiographs. False-positive rates have been reduced with use of computerized routines in digital mammography (32). Conversely, some investigators caution against relying on automated detection routines in direct digital mammography. Karssemeijer and Hendriks (33) found that the programs were accurate for detecting microcalcifications and stellate lesions on mammograms; however, performance was poorer if the targets were masses and asymmetric lesions. The authors cautioned that a large number of false-positive findings from the computer programs may be unmanageable in clinical practice.

An alternative to the CAD systems that alert a human observer is the use of computer-based classification systems. For example, one group reported success with a computer program designed to detect microcalcifications on mammograms, reporting a true-positive classification rate of 100% and a false-positive rate of 8% (34). In dental radiology, a machine classifier was accurate for differentiating between images of normal and diseased bone surrounding teeth (35–37). The gaze locations selected by observers were used as inputs to the learning routine. A similar recommendation was made for the interpretation of digital chest radiographs where observers' eye positions would identify regions of interest for localized image processing (38). Another recommendation was made to compare the performance of computer detection schemes with that of an observer while eye position is monitored (39). Mello-Thoms et al (40) developed a computer-assisted perception routine in which a computer program surrogate was used to predict the classification of observer decisions in mammography.

We performed this study to evaluate the performance of a computer-based classification system that uses the gaze locations of observers to define the subspace for machine learning. It was proposed that a computer-based learning system in which basic elements of the human visual system are incorporated can emulate human classification of novel cases. The elements included acquisition of essential information by means of fixation, formation of descriptions derived from primitives of gaze-selected sites, and classification through storage and retrieval of exemplar descriptions. The hypothesis was tested by comparing the performance of the computer-based system to that of an experienced observer as measured with tests of interobserver agreement.

We wanted to answer the following research questions: *(a)* How well does the machine classifier compare with an expert? and *(b)* Was the machine classification routine that used a gaze-selected sample space superior to that with a random selection of image space? To address these questions, three techniques were developed to sample image space, and the performance of the machine classifier was compared with the three techniques. To complete this task, eye position was recorded, the accuracies of the three techniques were assessed, and the classifications of the techniques were compared by using κ analyses (40).

MATERIALS AND METHODS

Images

We selected 32 dental radiographs from the patient archives at ACTA (Academic Center for Dentistry, Amsterdam, the Netherlands) that represented four categories of disease states in the periapical region (Fig 1): *(a)* no disease (normal tooth), *(b)* mild disease (widened periodontal ligament space [the radiolucent region that separates the root of the tooth from the lamina dura]), *(c)* moderate disease (destruction of the lamina dura), and *(d)* severe disease (resorption of bone in the periapical area). In some cases, there are only subtle differences between adjacent stages, and at times clinicians may only make a distinction in diagnosis when a variation in treatment would result. The nature of the disease is a pattern evidenced as the interface between the root of the tooth and the supporting bone. Therefore, the judgment is not of a focal lesion, but of disease presence over an area of the image.

Common image dimensions for chest images in eye motion studies range from 25 to 40 cm (41). This image size enables the observers to view the images at the same size as that used in a clinical setting. Intraoral dental radiographs measure 3 $\times$ 4 cm and are viewed at a distance of 20–30 cm. Therefore, the display of these images for eye motion studies necessitates substantial enlargement. The effect of this scaling has not been studied directly, although the results of the eye movement studies in dental radiology were consistent with those in which chest radiographs and mammograms were used (28). Dental radiographs were digitized as 256 $\times$ 256-pixel images at a depth of 8 bits. Images were enlarged to 512 $\times$ 512 pixels by using pixel replication and displayed as 20 $\times$

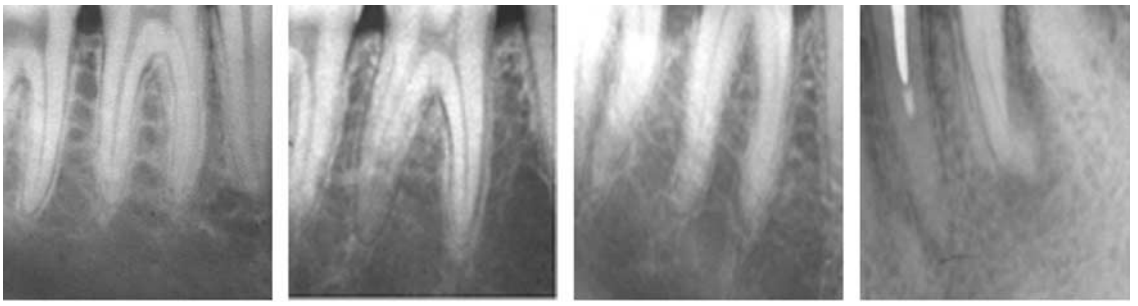

Figure 1. Examples of concept images. Radiographs show **(a)** normal periapical area, **(b)** mild disease (ie, widened periodontal ligament space, which is the radiolucent region that separates the root of the tooth from the lamina dura), **(c)** moderate disease (ie, destruction of the lamina dura), and **(d)** severe disease (ie, resorption of bone).

20-cm images on a 30 × 40-cm color video monitor (Trinitron; Sony, Ichinomiya, Japan) while eye motion was being recorded. The crown of the tooth was cropped from the digitized image so as to not give the observer a clue about the existing treatment of the tooth. At a viewing distance of 100 cm (the distance recommended by the equipment manufacturer), the displayed images subtended 11.5° × 11.5°. The resolution of the imaging system (camera, frame grabber) and the size of the original radiograph led to a displayed image in which 1 pixel represented 30 μm of the original object, a resolution sufficient to show the smallest details of interest (bone trabeculae are larger than 100 μm).

Observers

Seven staff members of the Academic Center for Dentistry (Amsterdam, the Netherlands) served as volunteers in two phases of the study. One individual (P.F.v.d.S.) with 24 years of experience in dental radiology served as the expert in phase I and classified archived dental radiographs to develop the sets of test images. He provided statements that served as the standard of reference. The expert did not participate in the eye-motion recording phase. The eye motion of six participants was recorded in phase II; these participants were trained in the Netherlands, Great Britain, or Greece. Two were radiologists with 15–18 years of experience in dental image interpretation; two were residents in their 3rd and 4th year of training, respectively; and two were dental students. All participants had 20/30 vision or better and sat for two 1-hour sessions while their eye position was monitored as they inspected images.

Eye-Position Recording and Analysis

The ISCAN eye tracker (ISCAN, Cambridge, Mass) was used to track eye position and collect information (42). Complete details of the procedures to calibrate and record eye position were reported elsewhere (35). ISCAN uses infrared illumination of a single eye to track the eye position. Two computers controlled image display and recording, and a small monitor displayed the eye during calibration and testing. Data were collected at 60 x-y pairs per second. The measured spatial resolution of the ISCAN used in these experiments was approximately 0.75° of visual angle, and the temporal resolution was 17 msec. Calibrations of the gray levels of the monitor were taken before each recording session by using an SMPTE, or Society of Motion Picture Test Engineers, test pattern.

Participants sat 100 cm in front of the monitor where images were displayed. A stationary chin rest was used to stabilize the position of the head during calibration and testing. Their task was to view the images and classify them into one of the four categories of disease. Before each trial, a calibration of eye position was performed to ensure spatial accuracy of recordings. For the calibration procedure, participants were required to look at the nine small circles displayed in a square pattern over the area subtended by the images. Trials lasted 16 seconds, during which eye position was recorded at 60 Hz. When participants classified the image, they looked to the side of the display and terminated search on that image. At the end of trials, the participant gave a hand signal to indicate the classification choice. Sessions of 16 trials lasted approximately 1 hour. A total of 192 records were collected (32 images × six observers). Calibration was performed and eye-position data were obtained with room lights off.

Eye-position data were spatially adjusted by using calibration information. Individual locations of corrected gaze were grouped into fixations and movements that were characterized by spatial and temporal constraints. Fixations were defined as a minimum of four consecutive

Table 1
Comparison of Observer and Expert Classifications

Observers' Classifications	Expert's Classification				
	Normal Tooth	Mild Disease	Moderate Disease	Severe Disease	Total
Normal tooth	35	7	4	0	46
Mild disease	8	20	12	0	40
Moderate disease	5	12	14	6	37
Severe disease	0	9	18	42	69
Total	48	48	48	48	192

Note.—Data are numbers of images. $\kappa = (35 + 20 + 14 + 42 - 48)/192 - (192/4) = 0.44$. The standard error of κ was 0.0464 (see reference 45 for a tutorial on the calculation of the standard error for κ). z test for κ, $z = .44/.0464 = 9.418$ ($P < .001$). Accuracy = ([35 + 20 + 14 + 42]/192) = 111/192 = 57.8%.

pairs of x-y points within 0.75° of visual angle for a minimum duration of 64 msec.

Machine Learning

Three techniques of sampling image space were developed: gaze-selected space, random selection, and constrained random selection. For the gaze-selected space, the regions of interest defined by means of recorded gaze were used as the image areas submitted to the learning algorithm. The eye-gaze unit for sampling was an individual fixation of 0.75° diameter. For the random space technique, a sample of fixations with a 0.75° diameter was generated with centers uniformly distributed in the image plane. The number of random samples per image was identical to the number of observer fixations. The third technique was similar to the random space technique except that the selection was constrained to central portions of the images by limiting the centers of "fixations" to lie within a 3° border interior to the image periphery. The rationale for the constrained border is that to maximize image processing during machine learning, samples were taken from central image space where the root of the tooth was shown. The constrained technique models an observer maximizing visual search to the central image space. Thus, each technique used the identical number of spatial samples; only the locations of the samples differed among the techniques.

All fixations placed by each observer on individual images were used as inputs to the classification system. The number of fixations varied for each observer viewing each image. The number of fixations individual observers placed on each image was used to determine the number of location samples for the two random methods.

In the learning phase, the classifier used a leave-one-out method, which is known to yield the best estimate of classifier performance (43). The leave-one-out technique enables classification of each of the 192 images, one at a time. The learning procedure did not assume that the input data were independent. Within a learning trial, the machine classifier learned on 191 samples and was tested on the 192nd sample. For all trials (six observers and 32 images), the samples were learned in random order. A total of 768 classifications (192 eye-movement records and four repetitions) were made by the machine for each technique (three techniques and 768 classifications for 2,304 runs). The learning strategy was incremental, and, therefore, the orders in which the 191 samples were presented helped determine, in part, the learned description. In the classification phase, the order may affect the accuracy of classifying the 192nd image and, thus, four repetitions through the data set were conducted to determine an overall level of accuracy.

Statistical Analyses

Performance was assessed by means of κ analyses (44). Observer classifications of the images were compared with the expert classification as a means of assessing observer performance. The relative performance of the three techniques was estimated by comparing machine classification with expert classification by using κ values and analyzed with nonparametric Kruskal-Wallis tests.

RESULTS

Observer Performance

The classification performance of the six observers was compared with the standard of reference of expert classification, as shown in Table 1. Overall accuracy for the 192 classifications was 57.8%, and the overall κ was 0.44.

The number of fixations placed on images before classification decisions were made ranged from 16 to 45. Estimates of the image area covered by a series of 0.75° diameter gaze locations range from 6% (for 16 fixations) to 18% (for 45 fixations), with an average of 12%. It is important to note that this is an estimate for the input to the machine classifier and not an estimate of the image areas used by the observer in classification. Figure 2 shows an example of a fixation pattern for one observer viewing an image that shows evidence of disease.

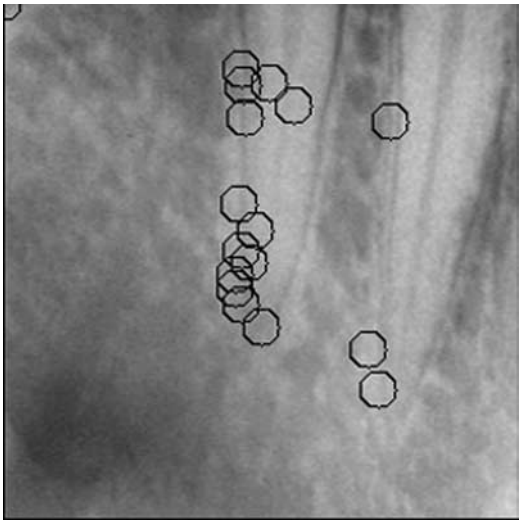

Figure 2. Example of fixation locations on a radiograph that shows widened ligament space. A circle with a diameter of 0.75 visual degrees represents each fixation. (Reprinted, with permission, from reference 35.)

Machine Classification

Performance of the machine classifier was calculated by comparing the classifications of images by the machine to that by the expert. For the gaze-selected technique, overall accuracy for the 192 images with the four trials was 83.4%, with a κ value of 0.780 (range of κ for four trials, 0.76–0.81; standard deviation [SD], 0.022) as shown in Table 2. Observers were clearly inferior to the machine classifier that used gaze-selected space, even though the observers and machine used the same image space information (Table 3).

Machine classification of the images with the three techniques was compared with the expert standard of reference by means of κ values. Accuracy for the random technique was 75.6%, with a κ value of 0.675 (SD, 0.025), and that for the constrained random technique was 76.9%, with a κ value of 0.692 (SD, 0.017). Results of a series of Kruskal-Wallis tests showed that gaze-selected space was superior to both the random ($\chi_2^2 = 5.398$, $P = .020$) and constrained techniques ($\chi_2^2 = 5.333$, $P = .021$). There were no differences in κ values between the random and constrained techniques ($\chi_2^2 = 0.337$, $P = .561$). Figure 3 illustrates the superior performance of the gaze-selected technique.

DISCUSSION

The purpose of the study was to evaluate the performance of a computer-based classification system that used the gaze locations of observers to define the subspace of the image for processing. We wanted to answer the following research questions: *(a)* How well does the machine classifier compare with an expert? and *(b)* Was the machine classification routine superior to the random selection of image space? To address these questions, three sampling techniques were developed to select image space, and the performance of the machine classifier was compared with the three techniques.

The machine classified dental images with 83% accuracy by using only a sample of an average of 17% of the image space. The accuracy of the machine was superior to that of the individual observers with use of the same selection of features and sample space. In addition, the machine classifier that used gaze-selected image areas was superior to that based on randomly selected areas. The classifier developed the schema required to classify images as showing specific disease patterns.

There are several situations in which a computer classification system for the diagnosis of dental disease would be helpful, including teaching novices, improving effectiveness and efficiency of diagnoses by experienced radiologists, and confirming diagnoses.

Table 2
κ Values for Machine versus Expert Classifications

	Trial				
Observer	1	2	3	4	Average
1	0.750	0.750	0.833	0.875	0.802
2	0.750	0.667	0.667	0.792	0.719
3	0.875	0.917	0.917	0.875	0.896
4	0.792	0.708	0.708	0.792	0.750
5	0.625	0.750	0.708	0.750	0.708
6	0.792	0.833	0.833	0.750	0.802
Average	0.764	0.771	0.778	0.806	0.780

Note.—Each trial consisted of 32 classifications per observer. The machine used gaze-selected sample space.

Table 3
Common κ Values for Observers versus Machine Classifier

Technique	Mean κ Value	95% Confidence Interval
Observer's gaze selected	0.44	0.185, 0.690
Machine trial		
1	0.764	0.558, 1.108
2	0.771	0.517, 1.026
3	0.778	0.523, 1.032
4	0.806	0.552, 1.061

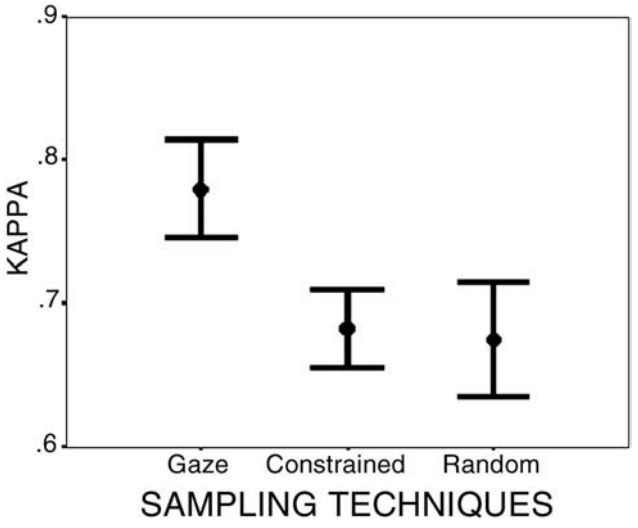

Figure 3. Graph illustrates the mean κ values and 95% confidence intervals estimated with the three sampling techniques.

There were two tasks modeled in this study. First, the selective interrogation of image space was examined. Second, the accuracy of the integration of selected features to form a classification of disease process was measured. Observers were efficient at selective interrogation of image space, as demonstrated by the superior performance of the machine classifier when using gaze-selected samples compared with randomly selected samples. Observers selected feature sets that are sufficiently rich to yield 83% accuracy by a machine classifier. This confirms the abilities of observers to select informative image samples, as has been previously demonstrated in visual search studies of chest radiographs (8–11), mammograms (18–21), and cross-sectional images (25–27). By using the same image samples as the observers, however, the performance of the machine classifier was superior to that of the observers. In addition, the machine was superior to observers when using randomly selected sample space. Perhaps observers do not fully utilize the visual information sampled during visual search. The findings suggest that when designing automated diagnostic systems, the entire image does not need to be interrogated for high accuracy.

It remains to be determined whether the selection process is improved by limiting the selection to areas based on the variables of the size of the region of interest and the duration of gaze. Perhaps the classification accuracy would improve by weighting the selected areas by the gaze durations and eliminating regions sampled for brief durations. The cognitive demands of the task and the level of expertise of the observer appear to drive the selection of the image areas by visual gaze. Improvements in classification accuracy may be achieved by selecting only experienced observers rather than including observers with limited experience. It may be possible to develop a standard set of image locations relative to the root of the tooth that would be identified by a technician who then would select the sample of the image for interrogation. This technique would eliminate the need to record eye position of experienced observers and enable rapid image processing and classification.

REFERENCES

1. Mol A. Image processing tools for dental applications. Dent Clin North Am 2000; 44:299–318.
2. White SC, Yoon DC, Tetradis S. Digital radiography in dentistry: what it should do for you. J Calif Dent Assoc 1999; 27:942–952.
3. van der Stelt PF. Principles of digital imaging. Dent Clin North Am 2000; 44:237–248.
4. Tuddenham WJ, Calvert WP. Visual search patterns in roentgen diagnosis. Radiology 1961; 76:255–256.
5. Tuddenham WJ. Visual search, image organization, and reader error in roentgen diagnosis. Radiology 1962; 78:694–704.
6. Thomas EL, Lansdown EL. Visual search patterns of radiologists in training. Radiology 1963; 81:288–291.
7. Thomas EL. Search behavior. Radiol Clin North Am 1969; 7:403–417.
8. Kundel HL, Wright DJ. The influence of prior knowledge on visual search strategies during the viewing of chest radiographs. Radiology 1969; 93:315–320.
9. Kundel HL, LaFollette PS. Visual search patterns and experience with radiological images. Radiology 1972; 103:523–526.
10. Kundel HL, Nodine CF, Carmody D. Visual scanning, pattern recognition and decision-making in pulmonary nodule detection. Invest Radiol 1978; 13:175–181.
11. Kundel HL, Nodine CF, Toto LC. Searching for lung nodules: the guidance of visual search. Invest Radiol 1991; 26:777–781.
12. Carmody DP. Free search, restricted search, and the need for context in radiologic image perception. In: Groner R, McConkie G, Menz C, eds. Eye movements and human information processing. Amsterdam, the Netherlands: Elsevier Science, 1985; 341–355.
13. Carmody DP, Nodine CF, Kundel HL. Global and segmented search for lung nodules of different edge gradients. Invest Radiol 1980; 15: 224–233.
14. Carmody DP. Lung tumour identification: decision-making and comparative scanning. In: Gale AG, Johnson FS, eds. Theoretical and applied aspects of eye movement research. Amsterdam, the Netherlands: North Holland, 1984; 305–312.
15. Carmody DP, Nodine CF, Kundel HL. Finding lung nodules with and without comparative visual scanning. Percept Psychophys 1981; 29: 594–598.
16. Carmody DP, Kundel HL, Toto LC. Comparison scans while reading chest images: taught, but not practiced. Invest Radiol 1984; 19:462–466.
17. Nodine CF, Kundel HL, Polikoff JB, Toto LC. Using eye movements to study decision making of radiologists. In: Luer G, Lass U, Shallo-Hoffman J, eds. Eye movement research: physiological and psychological aspects. Gottingen, Germany: Hogrefe, 1988; 349–363.
18. Barrett JR, deParedes ES, Dwyer SJ III, Merickel MB, Hutchinson TE. Unobtrusively tracking eye gaze direction and pupil diameter of mammographers. Acad Radiol 1994; 1:40–45.
19. Gale AG, Savage CJ, Pawley EF, Wilson AR, Roebuck EJ. Breast screening: visual search and observer performance. SPIE Med Imaging Proc 1994; 2166:66–75.
20. Krupinski EA. Visual scanning patterns of radiologists searching mammograms. Acad Radiol 1996; 3:137–144.

21. Nodine CF, Kundel HL, Lauver SC, Toto LC. The nature of expertise in searching mammograms for breast lesions. SPIE Med Imaging Proc 1996; 2712:89–94.
22. Berbaum KS, Franken EA Jr, Dorfman DD, et al. Cause of satisfaction of search effects in contrast studies of the abdomen. Acad Radiol 1996; 3:815–826.
23. Hu CH, Kundel HL, Nodine CF, Krupinski EA, Toto LC. Searching for bone fractures: a comparison with pulmonary nodule search. Acad Radiol 1994; 1:25–32.
24. Berbaum KS, Brandser EA, Franken EA, Dorfman DD, Caldwell RT, Krupinski EA. Gaze dwell times on acute trauma injuries missed because of satisfaction of search. Acad Radiol 2001; 8:304–314.
25. Niimi R, Shimamoto K, Sawaki A, et al. Eye-tracking device comparisons of three methods of magnetic resonance image series displays. J Digit Imaging 1997; 10:147–151.
26. Beard DV, Pisano ED, Denelsbeck KM, Johnston RE. Eye movement during computed tomography interpretation: eyetracker results and image display-time implications. J Digit Imaging 1994; 7:189–192.
27. Beard DV, Johnston RE, Toki O, Wilcox C. A study of radiologists viewing multiple computed tomography examinations using an eye-tracking device. J Digit Imaging 1990; 3:230–237.
28. Schouten E, van der Stelt PF. Expertise in interpreting dental radiographs. III. Bioengineering in dentistry. In: Proceedings of the 14th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Piscataway, NJ: IEEE Engineering in Medicine and Biology Society, 1992; 1105–1106.
29. Krupinski EA, Nodine CF, Kundel HL. Perceptual enhancement of tumor targets in chest x-ray images. Percept Psychophys 1993; 53:519–526.
30. Giger ML, Doi K, MacMahon H, Metz CE, Yin FF. Pulmonary nodules: computer-aided detection in digital chest images. RadioGraphics 1990; 10:41–51.
31. MacMahon H, Engelmann R, Behlen FM, et al. Computer-aided diagnosis of pulmonary nodules: results of a large-scale observer test. Radiology 1999; 213:723–726.
32. Ema T, Doi K, Nishikawa RM, Jiang Y, Papaioannou J. Image feature analysis and computer-aided diagnosis in mammography: reduction of false-positive clustered microcalcifications using local edge-gradient analysis. Med Phys 1995; 22:161–169.
33. Karssemeijer N, Hendriks JH. Computer-assisted reading of mammograms. Eur Radiol 1997; 7:743–748.
34. Davies DH, Dance DR. Automatic computer detection of clustered calcifications in digital mammograms. Phys Med Biol 1990; 35:1111–1118.
35. Kerbaugh McGrath SP. Concept learning with gaze selected sample space. Dissertation. Rutgers University, Piscataway, NJ, 1996.
36. McGrath SP, Dunn SM, Kantor ML, Carmody DP, van der Stelt-Schouten E, van der Stelt PF. Image processing with gaze selected sample space. Presented at the Eye Movements and Visual Search in Dentistry Symposium held at the Sixth “Far West” Image Perception Conference, Philadelphia, Pa, October 13–15, 1995.
37. McGrath SP, Dunn SM, Carmody DP, Kantor ML, van der Stelt-Schouten E, van der Stelt PF. Gaze selected sample space for machine learning. Dentomaxillofac Radiol 1997; 26:275–276.
38. Barrett JR. Eye movements can control localized image enhancement and analysis. RadioGraphics 1997; 17:525–530.
39. Krupinski EA, Nishikawa RM. Comparison of eye position versus computer identified microcalcification clusters on mammograms. Med Phys 1997; 24:17–23.
40. Mello-Thoms C, Dunn SM, Nodine CF, Kundel HL. An analysis of perceptual errors in reading mammograms using quasi-local frequency spectra. J Digit Imaging (in press).
41. Kundel HL, Nodine CF, Thickman D, Toto L. Searching for lung nodules: a comparison of human performance with random and systematic scanning models. Invest Radiol 1987; 22:417–422.
42. Razdan R, Kielar A. Eye tracking for man/machine interfaces. Sensors 1988; 5:39–45.
43. Kulikowski CA, Weiss S. Computer systems that learn: classification and prediction methods from statistics, neural nets, machine learning, and expert systems. San Francisco, Calif: Morgan Kaufmann, 1991.
44. Fleiss JL. Statistical methods for rates and proportions. 2nd ed. New York, NY: Wiley, 1981; 212–236.
45. Herman PG, Khan A, Kallman CE, Rojas KA, Carmody DP, Bodenheimer MM. Limited correlation of left ventricular end-diastolic pressure with radiographic assessment of pulmonary hemodynamics. Radiology 1990; 174:721–724.