Computer Diagnosis of Pneumoconiosis

RICHARD P. KRUGER, MEMBER, IEEE, WILLIAM B. THOMPSON, MEMBER, IEEE, AND A. FRANKLIN TURNER

Abstract—**The advent of increased government involvement in occupational health maintenance with compensation for affected individuals is requiring new approaches to medical decision making. One aspect of this involvement will perhaps include the automatic mass diagnostic screening of medical films for the detection of a specific abnormality with an occupational etiology. The results of two complementary approaches for performing diagnostic screening for presence and profusion of coal worker's pneumoconiosis from the routine posterior–anterior chest radiograph are presented. The first is a digital approach utilizing the measurement of image texture, while the second uses hybrid optical–digital methods involving the optical Fourier transform. Both approaches yielded classification results comparable to experienced radiologists.**

INTRODUCTION

PUBLIC LAW 91173 has specified that each coal worker has the right to obtain his chest roentgenogram at regular intervals not to exceed 5 years to detect the possible presence and progression of coal worker's pneumoconiosis (CWP). There are approximately 130 000 active miners who are covered by this act. The Social Security Administration has also been reviewing approximately 200 000 claims for disability due to CWP from active or retired coal workers or their survivors. Each roentgenogram is read by at least two certified readers according to the ILO U/C 1971 international classification of radiographs of the pneumoconioses [1]–[3]. This semiquantitative description of the radiographic changes due to the pneumoconioses is designed to provide both clinical and epidemiological information. From a clinical and occupational health viewpoint, the classification of radiopacities as to *profusion* category is of primary importance. Profusion of small regular and irregular opacities is judged on 12 minor and 4 major categories based on the number of opacities per unit of lung area. A fifth major profusion category consists of still larger regularly and irregularly shaped lesions associated with progressive massive fibrosis (PMF). This fifth category often occurs when smaller lesions coalesce. These five categories are shown in Table I. Predominant small lesion *type* is also indicated with regular (rounded) opacities indicated by the letters *p*, *q*, and *r* and irregular opacities by the letters *s*, *t*, and *u* with respect to increasing size. However, this latter information is of more epidemiological than clinical or compensative value.

Manuscript received February 12, 1973; revised July 9, 1973. This work was supported by the National Institute for Occupational Safety and Health under Contract HSM-99-72-30, and by the Advanced Research Projects Agency of the Department of Defense monitored by the Air Force Eastern Test Range under Contract F086-06-72-C-0008.

R. P. Kruger is with the Department of Electrical Engineering, University of Southern California, Los Angeles, Calif. 90007.

W. B. Thompson is with the Computer Science Program, University of Southern California, Los Angeles, Calif. 90007.

A. F. Turner is with the Department of Medicine and Radiology, University of Southern California, Los Angeles, Calif. 90007.

TABLE I
PROFUSION CATEGORIES AND PROBABILITY OF OCCURRENCE

Major profusion category	Corresponding minor profusion categories	Number of cases	A priori class probabilities
0 (Simple	0/- 0/0 0/1	51,491	.878
1 disease)	1/0 1/1 1/2	4,630	.079
2	2/1 2/2 2/3	1,772	.030
3	3/2 3/3 3/4	138	.020
4 (Complicated disease)	PMF	688	.011

Category 0 Normal or doubtful
1 Opacities definitely present but few in number
2 Opacities numerous but normal lung vascularity still visible
3 Opacities very numerous with lung vascularity obscured
4 Evidence of coalescence (PMF)

To determine overall film profusion category, a decision as to the *extent* of the opacities in the lung field must first be ascertained. Each lung field is divided into apical, medial, and basal zones by horizontal lines drawn at one-third and two-thirds of the vertical distance between the apex of the lung and the dome of the diaphragm. Overall film profusion category is established by visually averaging the profusion in the most affected zones. The result is a visually arrived at indication of zonal extent, opacity type, and profusion category for each film. Fig. 1 yields a visual comparison of a normal right apical zone with three other diseased apical zones. Recent investigations [4], [5] seem to support the contention that, for simple pneumoconiosis, the 12-point ILO U/C profusion categories from 0/–3/4 exhibit an approximate linear relationship to the amount of dust deposition in the lungs.

Current manual screening utilizing the aforementioned international standards consists of a hierarchy of readers with increasing diagnostic skills relative to this disease. The initial reader, designated an A reader, is either a radiologist or a general practitioner who is located in the region where the radiograph was obtained. The second or B reader is often geographically remote and is a board certified radiologist at one of a small number of pneumoconiosis screening centers established nationwide. If there is more than one minor profusion category difference between the A and B readings, a *significant disagreement* exists, and one of the C readers decides a final category. In rare instances a panel of C readers constitute a D reader. There are approximately 150 A readers, 30 B readers, and 5 C readers actively screening for this disease. It should also be noted that the 5 C readers are largely responsible

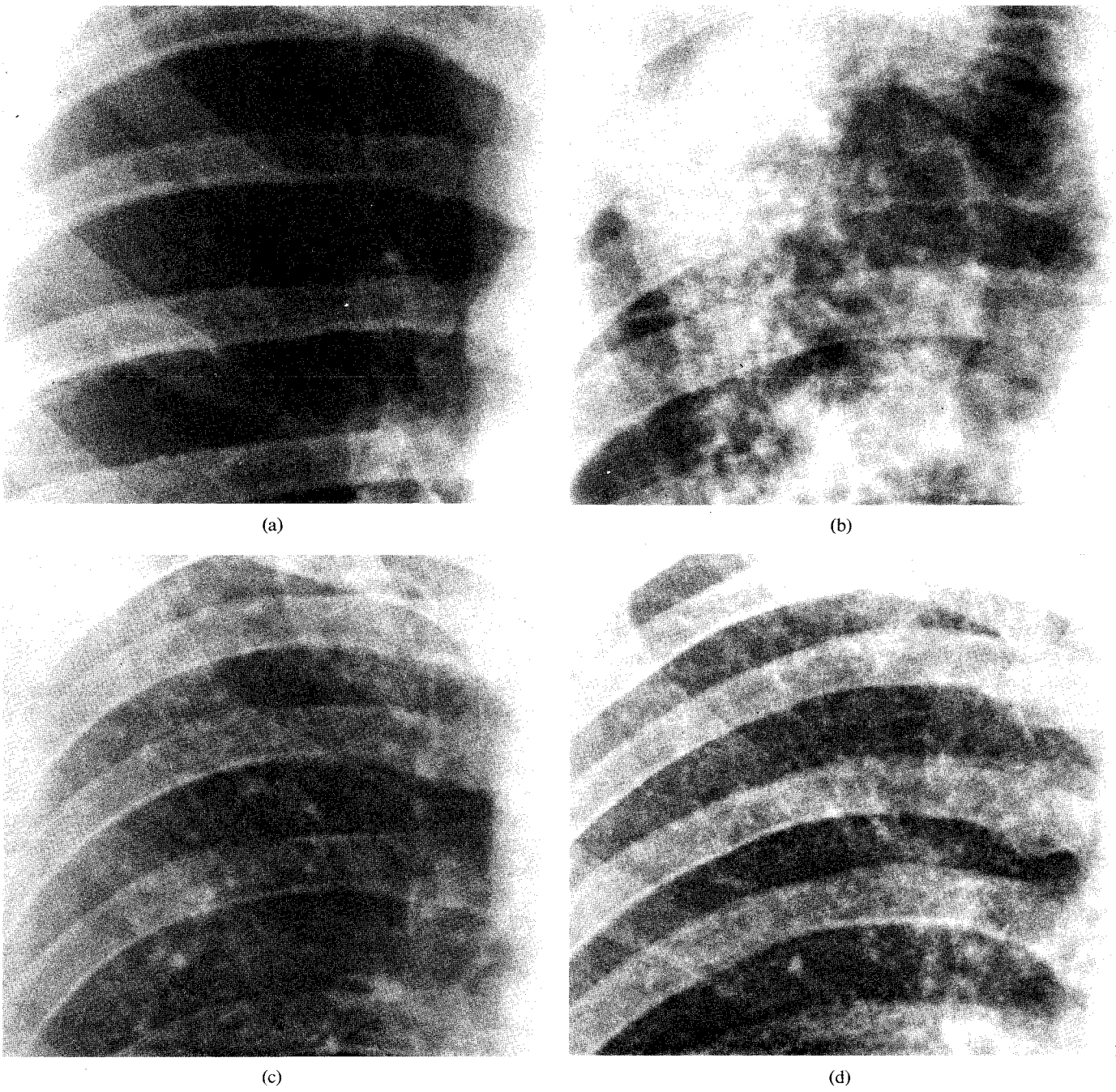

Fig. 1. Normal and several abnormal lung zones. (a) Normal. (b) *u* 2/2 *B*. (c) *t* 1/2. (d) *p* 2/2.

for the aforementioned ILO U/C standards. Recently published NIOSH epidemiological data [6] from this reader hierarchy is shown in Table I. It indicates that approximately 13 percent of active miners are diseased. It was also reported that perhaps 28 percent of retired miners are diseased. The reported significant difference in radiological category between A and B readers which necessitated a C reading was 27 percent. This interreader variation in pneumoconiosis screening is not surprising when viewed within the context of earlier studies of intra- and interreader variation in chest radiographic diagnosis of similar-appearing tuberculosis lesions conducted by Garland [7] and Yerushalmy [8].

Reger and Morgan [9], [10] also conducted studies concerning the effects of interreader variation and film quality on assignment of profusion category. In only 56.7 percent of the cases was there agreement on a four diagnostic class assignment based on increasing profusion category among four experienced readers of pneumoconiosis films. The percentage was surprisingly low since approximately one-half of the films selected showed no evidence of pneumoconiosis, although many were suffering from other respiratory ailments such as tuberculosis or showed evidence of other radiographic abnormalities. It was also found that film quality had an appreciable and significant effect on diagnostic categorization. A marked tendency to place over-

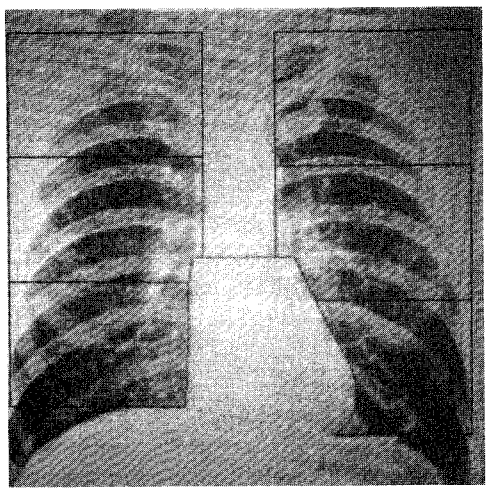

Fig. 2. Chest radiograph with superimposed computer-calculated cardiac outline and lung zones.

exposed films in a lower profusion category than the same films properly exposed was noted. It was concluded that standardized exposure, suggested by Jacobson [11], and refusal to interpret over- or under-exposed films would be advantageous. Finally, intrareader variations on several examinations of the same films with experienced readers was less significant than the other previously mentioned effects.

Based on present projections of future reading loads, it is likely that present manual reading procedures will result in a progressive overburdening of the radiographic diagnostic system as it currently exists. Thus some automated procedure for diagnostic mass screening of such films will be desirable and most probably necessary. Since it has been demonstrated that a high percentage of all films read are diagnosed as not showing radiographic evidence of pneumoconiosis, a cost effective and rapid automatic film reader to consistently screen out all definite normals would significantly relieve the projected overburdening previously referred to. In addition, it is also conceivable that this automatic system could produce quantitative indexes pertinent to the extent of losses of normal pulmonary vascular patterns by irregular *s*, *t*, and *u* type and rounded *p*, *q*, and *r* type radiopacities. These quantitative measurements would be quite consistent in time. There would therefore be little intrareader variation since fixed quantitative measurements are minimally subject to variations between serial examinations of the same individual.

It is fortunate that there presently exists the semiquantitative grading procedure for radiographic diagnosis of the pneumoconioses. This type of descriptive procedure for diagnosis is usually not available. It has provided a framework within which the physicians and technologists can communicate in undertaking a study of possible means by which this diagnosis can be automated.

Previous research has indicated that automatic anatomical feature location, measurement, and diagnostic classification of the superior mediastinal and cardiac projections in large numbers of standard posterior–anterior chest radiographs is feasible with accuracy rates comparable to manual diagnosis [12], [13]. A modified version of that algorithm will also divide the right and left lung fields into six rectangular zones as stipulated in the ILO U/C grading procedure. It would therefore seem possible to automatically locate each lung zone for subsequent machine diagnosis. It is also expected that existing cardiac diagnosis capabilities will be useful in detecting suspect cor pulmonale, abnormal cardiac size and shape, and possibly ill-defined cardiac outline as stipulated in the existing physician grading procedure. An example of a computer-detected cardiac boundary and lung zone determination is shown in Fig. 2.

Image Textural Discrimination

The problem of manually detecting and grading simple pneumoconioses opacities from radiographs appears to be largely one of discrimination between normal pulmonary vascularity (lung marking) patterns and partial or complete obliteration of this normal tree-like structure by opacities of various sizes and profusions which themselves exhibit a more or less textural nature. Complicated disease, involving progressive massive fibrosis is characterized by larger radiopacities which perhaps are not characterized nearly so well by a textural description. Rosenfeld and Troy [14] described image texture ideally as the "repetitive arrangement of a unit pattern over a given area." It was also stated that, in natural imagery, it is often difficult to identify such unit patterns or determine their repetitive arrangement. Therefore, the previous description should be used only as a guide in the analysis of natural imagery.

Pickett [15] referred to two kinds of human subjective textural analyses. These might be described as deliberate and impressionistic. In the latter, for example, an immediate impression of the coarseness or fineness of a visual texture is received. In a sense, the observer gets an immediate answer to a set of questions not consciously asked. Deliberate analysis, on the other hand, implies that the observer is looking more closely at a pattern or texture, as would perhaps be the case in trained observation. It was also concluded that many textures are discriminated by assessments of size, shape, orientation, and repetition rate. In empirical studies Gibson [16] has shown that regular textures seem to convey stronger impressions than irregular ones, and he also concluded that his observers relied on impressionistic analysis. Therefore, psychovisually, a sensation described as texture is perceived by the human. However, this sensation is less well understood than visual contour identification. It can also be stated that corresponding computer analysis of image textures is not as well advanced as contour or boundary analysis. It may well be that the two will not be mutually exclusive concepts from either a visual or computer analysis viewpoint.

In computer analysis of natural image texture, Hawkins [17] made several observations. First, the notion that natural texture is strictly locally repetitive is only approximately true. That is, what is replicated is a pattern class within which all examples are regarded as equivalent. Furthermore, replication may be subject to spatial nonlinear phase shifts and, therefore, only be approximate. Second, nonrandom arrangements of pattern parts appear

to be associated with meaningful textures. He concluded with the observation that texture classification may be one of the more difficult tasks in the field of image processing.

Since the approach to textural feature extraction of pneumoconiosis opacities were necessarily experimental, some probable statistical indexes were explored. With statistical analysis it was hoped that useful measures of the organization and arrangement of the textures in question would be derived without having to focus attention on the specific structural properties. For instance, the chosen statistical measurements did not attempt to recognize or trace the normal vascular tree or the pneumoconiotic lesions per se, but merely to compute quantitative impressions which characterize them.

Textural discrimination of images was attempted in both the spatial and spatial frequency domains. Therefore, the specific approach was to consider pneumoconiotic opacities or normal lung markings as distinct image textures. The task was to experimentally compute several textural features and subsequently apply a feature selection and supervised statistical pattern recognition technique to obtain a diagnostic classification for discrimination between normal and abnormal lung zones.

The Film Data Base

It was decided to create a data base consisting of 4- by 5-in zonal reproductions. There were several reasons for this choice, not the least of which was the fact that neither the digital nor the optical device was capable of processing the complete 14- by 17-in film for input to the computer. There was also a conscious effort to accurately establish the profusion category and lesion type for purposes of computer training. It was felt that zone selection would be an aid toward this goal. The following specifications were followed on each zonal reproduction. Each reproduced lung zone was to be one to one in size with the equivalent region on the full film. The zonal selections were to provide an ample mix of all 6 zones to minimize zonal bias during computer training. An effort was made to select an equal number of zonal films from the 12 minor profusion categories, 2 lesion types, and 6 lesion sizes characteristic of simple pneumoconiosis. These criteria were an effort to create a data base with minimal zonal bias and a uniform distribution of lesion profusion categories, types, and sizes.

The problem under consideration was as follows: having demonstrated the computer detectability of lung zones, could a machine render a diagnosis within those zones competitive with manual diagnosis?

The film base for the feasibility study was selected by Drs. G. Jacobson and A. F. Turner of the Los Angeles County–University of Southern California Medical Center Department of Radiology. Of the 102 abnormal zonal films, approximately 40 percent were extracted from standard films used by the American College of Radiology (ACR) to train physicians to detect and grade pneumoconiotic lesions. The remaining abnormal zonal films were selected and graded by Drs. Jacobson and Turner as a committee. Thirty-three of the normal zonal films were selected from chest radiographs of LA County–USC employees with no known work history in a dusty environment and no clinical or radiographic evidence of lung dysfunction. Six zonal films from ACR standard films from profusion category 0/1 were also included. All films were judged of acceptable quality according to the ILO U/C classification criteria.

A representative sample which consisted of 95 zonal reproductions was digitized by a flatbed scanning microdensitometer at a resolution of 9.8 pixels/mm. The optical studies used all 141 reproductions by comparison.

Data Management and Textural Feature Extraction from the Digital Images

It was decided the visual diagnosis of simple pneumoconiosis lesions is often arrived at by inspecting the lung regions between the more visually prominent posterior ribs. This seemed logical since it is in these interrib spaces that normal vascularity is least obstructed by visual interference from more dense radiographic structures. Therefore, posterior interrib spaces were manually extracted in the following manner. First a computer-generated grid was superimposed upon each of the digitized images. This grid allowed visual extraction of from 4 to 10 squares from each interrib space. These squares are delineated for the zonal film shown in Fig. 3. When extracting squares, an effort was made to avoid the least radiographically dense portion of the hilar regions. However, no effort was made to avoid the less prominent anterior ribs or interrib spaces not involved with disease. Since there were 3 to 4 such interrib spaces per zonal film, the complete digital data base consisted of over 1800 such squares from 298 interrib spaces.

It was decided to consider *each* posterior interrib space as a *separate* input to the diagnostic classifier in order to create an automated classification technique not sensitive to any interrib space or any specific lung zone. This would constitute a *local* region for textural analysis. Thus the digital input data base consisted of 298 samples.

The data from *all* the squares in an individual interrib space were combined to form a gray-level histogram for that complete interrib space. A transformation of gray levels was performed which produced an interrib space with eight equally likely gray levels [18]. This preprocessing step was designed to negate the effect of monotonic distortions introduced due to inconsistency in photography and/or digitization of the original images by constraining all inputs to the feature extractor to be identical with respect to first-order probability of gray-level occurrence. The spatial textural measures were all based on spatial gray-level dependence matrices [19]–[21] under the assumption that visual texture-context information in an interrib space is contained in the spatial relationship between image picture elements at several fixed distances and angular orientations. More specifically, it was assumed that this texture-context information is adequately specified by the symmetric matrix of relative frequencies $p(i,j)$ with which two neighboring pixels are separated by a distance (d) and an angle (a) for each (i,j) gray-level pair in the space. For this application, d is the number of image lines separating the two pixels of interest. An 8×8 symmetric count matrix was formed within each square with all count matrices summed

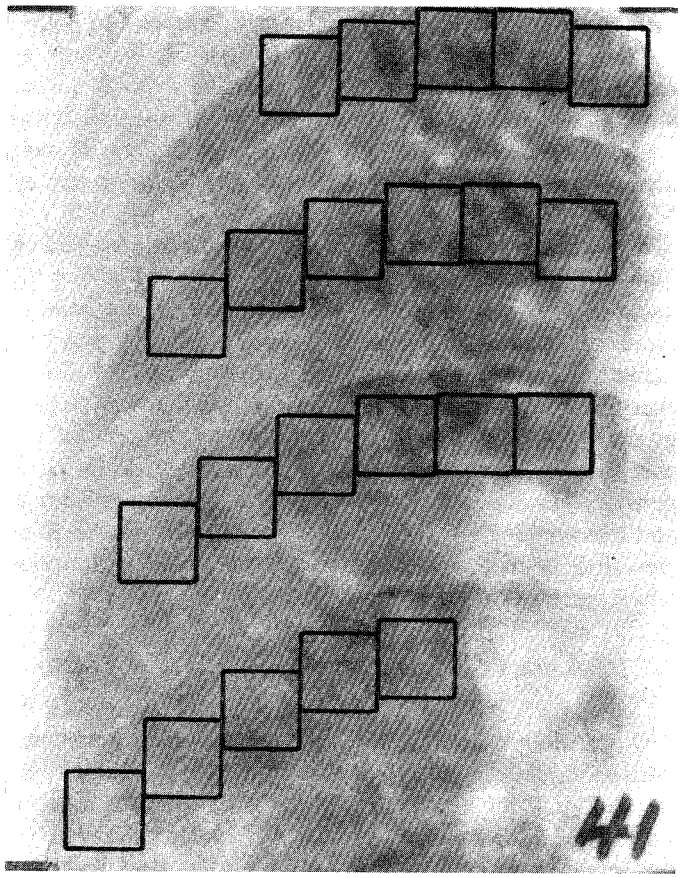

Fig. 3. Zonal film with extracted interrib space squares.

and normalized to create a matrix of relative frequencies for each interrib space as a function of a and d

$$p(i,j,a,d) \tag{1}$$

where

$$i = 0,1,\cdots,7, \qquad j = 0,1,\cdots,7,$$
$$a = 0°,45°,90°,135°, \qquad d = 1,3,7,11.$$

For each of the 298 spaces there were 16 such 8×8 matrices per space. The following 5 textural measurements $T_k(a,d)$, $k = 1,\cdots,5$ were computed for each matrix

$$T_1(a,d) = \sum_{i=0}^{7}\sum_{j=0}^{7} ijp(i,j,a,d) \tag{2}$$

$$T_2(a,d) = \sum_{i=0}^{7}\sum_{j=0}^{7} (i-j)^2 p(i,j,a,d) \tag{3}$$

$$T_3(a,d) = \sum_{i=0}^{7}\sum_{j=0}^{7} \frac{p(i,j,a,d)}{1+(i-j)^2} \tag{4}$$

$$T_4(a,d) = \sum_{i=0}^{7}\sum_{j=0}^{7} p(i,j,a,d) \log p(i,j,a,d) \tag{5}$$

$$T_5(a,d) = \sum_{i=0}^{7}\sum_{j=0}^{7} |i-j| p(i,j,a,d) \tag{6}$$

for $a = 0°$, 45°, 90°, 135°, and $d = 1,3,7,11$. T_1 is an autocorrelation measure designed to measure image coarseness. T_2 is a dissimilarity measure often called the moment of inertia. T_3 measures the extent to which the same or similar gray levels tend to be neighbors. T_4 is a conditional entropy measure and measures image homogeneity. T_5 is another dissimilarity measure which is similar to T_2. A total of 80 textural measures were extracted from each interrib space with each textural feature a function of angle a and distance d. The number of textural features was reduced from 80 to 60 by calculating the mean $\bar{M}$, variance V, and range R at a given distance d for each of 4 angles a

$$\bar{M}_k(d) = \frac{1}{4}\sum_{a=1}^{4} T_k(a,d) \tag{7}$$

$$R_k(d) = \max T_k(a,d) - \min T_k(a,d) \tag{8}$$

$$V_k(d) = \frac{1}{4}\sum_{a=1}^{4} (T_k(a,d) - \bar{M}_k(d))^2 \tag{9}$$

for $k = 1,\cdots,5$, $d = 1,3,7,11$, and $a = 0°,45°,90°,135°$. The final $\bar{M}$, R, and V measures did not possess a strict directional bias and were therefore explicitly a function of d and only implicitly a function of a.

Fourier Transform Domain Feature Extraction Using A Coherent Optical Approach

The textural features extracted in the previous section were derived from digital spatial domain data localized to individual posterior interrib spaces. The Fourier domain measures treated each of the 141 lung region films as an entity and as such measured more *global* aspects of visual texture which included the influence of both posterior and anterior rib projections.

Fig. 4 describes in simplest form the Recognition Systems, Inc., ROSA-3 device used to extract the spatial frequency measures of visual texture. A helium–neon laser emitted a light which passed through a collimating lens and then through the input film image. The transmitted light from the film was next passed through a positive thin lens which performed the Fourier transformation. The transformed image was then projected onto a detector and appropriate energy measurements were obtained. For the images in this experiment, a circular 2.5-in-diameter aperture illuminated a circular area in the center of the appropriate lung zone film. Fig. 5 is an example of an illuminated circular area within a typical zonal film.

The Fourier transform equation is

$$\hat{F}(u,v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y) \exp\left[j2\pi(xu+yv)\right] dx\, dy \tag{10}$$

where $f(x,y)$ is the illuminated film region and $\hat{F}(u,v)$ is the transform of that region. The chosen transform textural measures were insensitive to phase and as such measured aspects of the magnitude of the transform denoted as $|\hat{F}(u,v)|$.

It is well known that high-frequency information pertains to the amount of edge information in an image. It can be hypothesized that lung regions with pneumoconiosis opacities will also generate more of this higher frequency edge information than a normal film. Several investigators [22]–[24] have previously used this transform property for terrain, cell, and lung vascularity classification studies with some degree of success. Therefore, a detector in the trans-

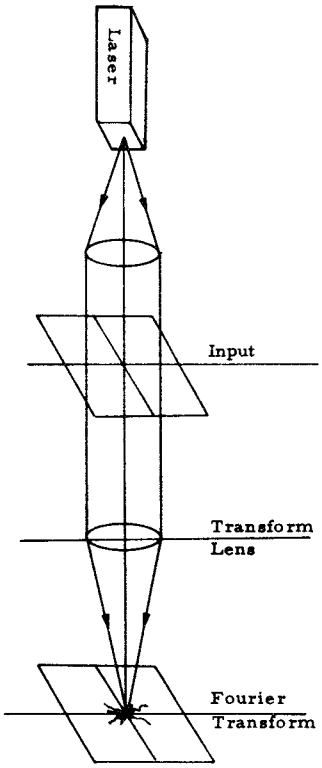

Fig. 4. Basic configuration of Recognition Systems, Inc., Rosa-3 optical spectrum analyzer.

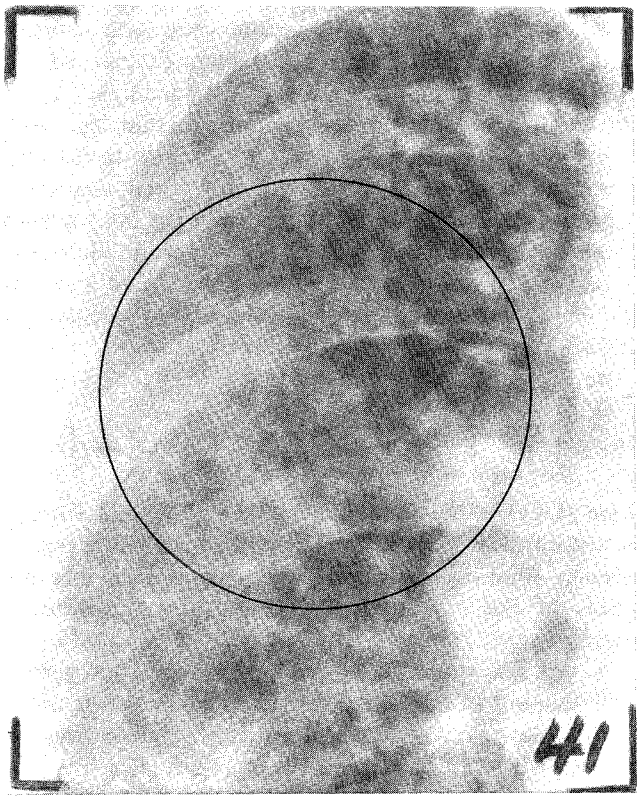

Fig. 5. Zonal film with 2.5-in aperture superimposed.

form plane consisting of 32 annular rings was used to obtain the total contribution of 32 radial frequencies to accurately reflect the amount of edge by the relative strength of the annulli. The maximum spatial frequency measurable in this experiment corresponded to 8.8 line pairs/mm. This exceeds observable resolution for this class of films as measured by Morgan [25] and Rossmann [26].

A vertically oriented 20° wedge-shaped band-rejection filter in the detector served both as a read-out path for the detected energy in the 32 annulli and also as a means to reduce the effect of the posterior ribs in the energy measurements. The spectral measurements were normalized to unit energy by dividing each measurement by the total energy in the transform, which in this case was the sum of the energy in all the annular rings. This normalization compensated for linear interfilm differences in dynamic range and average density. The normalized energies in the rings were then logarithmically transformed to create distributions which were more nearly Gaussian.

Textural Feature Selection and Classification Approach

Quite often a group of feature measurements contains both redundant features and those which are of little value in separating the classes. For a classifier to work successfully, these features must be removed. In addition, the larger the set of measurements the classifier must deal with, the greater the numerical inaccuracy in computing the needed discriminant functions. Thus it is advantageous to reduce the dimensionality of the feature space by determining which of the original measurements contain the most useful information for the classifier.

In order to select this subset, a measure of the value of a feature must be defined. For a two-class problem where the classifier assumes a Gaussian distribution of features, the "divergence" [27] measure is often used. The divergence of a set of features $\bar{x}$ is defined as

$$J(S_1,S_2) = \int_{-\infty}^{\infty} [p(\bar{x}) \mid S_1) - p(\bar{x} \mid S_2)] \ln \left[\frac{p(\bar{x} \mid S_1)}{p(\bar{x} \mid S_2)}\right] d\bar{x} \tag{11}$$

where $\bar{x}$ is an R-dimensional vector, S_1 denotes class 1, and S_2 denotes class 2. It can be shown that maximizing the divergence measure minimizes the bound on the probability of error P_e such that

$$P_e \leq (P(S_1)P(S_2)) \left[\frac{J(S_1,S_2)}{4}\right]^{-1/4} \tag{12}$$

when the classifier is a maximum-likelihood estimator and the underlying distributions are multivariate Gaussian.

If classes S_1 and S_2 are assumed to be multivariate Gaussian distributed, the R-dimensional $\bar{x}$ vector is distributed

$$p(\bar{x}/S_k) = \frac{\exp\left[-\frac{1}{2}(\bar{x} - \bar{\mu}_k)^t[\Phi_k]^{-1}(\bar{x} - \bar{\mu}_k)\right]}{(2\pi)^{R/2}|[\Phi_k]|^{1/2}} \tag{13}$$

$$p(x/S_k) \sim N(\bar{\mu}_k,[\Phi_k]), \qquad k = 1,2$$

where $\bar{\mu}_k$ is a mean vector and $[\Phi_k]$ is a covariance matrix for each class. The divergence distance is optimized over all linear transforms $[T]$ of dimension $N \times R$, which implies Gaussian distributions in the transformed space as well. Thus

$$p(\bar{x}/S_k[T]) \sim N(\bar{\mu}_k{}^T,[\Phi_k{}^T]), \qquad k = 1,2$$
$$\bar{\mu}_k{}^T = [T]\bar{\mu}_k, \qquad [\Phi_k{}^T] = [T]^t[\Phi_k][T]. \quad (14)$$

Therefore, the divergence measure becomes a function of the linear transformation and will be denoted $J(S_1,S_2,T)$. The divergence is therefore expressed as [27]

$$J(S_1,S_2,T) = \tfrac{1}{2}\,\mathrm{tr}\,([\Phi_2{}^T]^{-1}[\Phi_1{}^T] + [\Phi_1{}^T]^{-1}[\Phi_2{}^T] - 2[I]) + \tfrac{1}{2}\,\mathrm{tr}\,(([\Phi_1{}^T]^{-1} + [\Phi_2{}^T]^{-1}) \times ((\bar{\mu}_1{}^T - \bar{\mu}_2{}^T)(\bar{\mu}_1{}^T - \bar{\mu}_2{}^T)^t)) \quad (15)$$

where tr is the trace of the matrix.

It would be desirable to find which combination of N features, taken together, would be optimal. However, in practice this is often not feasible. A compromise is to calculate the divergence of R features one at a time and then to choose N of those with the highest divergence value. If Gaussian statistics are assumed and if features are analyzed one at a time, the divergence measure of the ith feature becomes

$$J(S_1,S_2,i) = \frac{(\sigma_1{}^{(i)} - \sigma_2{}^{(i)})^2 + (\sigma_1{}^{(i)} + \sigma_2{}^{(i)})(\mu_1{}^{(i)} - \mu_2{}^{(i)})^2}{2\sigma_1{}^i\sigma_2{}^i}, \qquad i = 1,\cdots,R \quad (16)$$

where $\sigma_k{}^i$ and $\mu_k{}^i$ are the variance and mean of feature i and class k, and $[T]$ is a $1 \times R$ matrix with one nonzero unity term at location $(1,i)$.

The divergence measure is defined for only a two-class situation. In order to use this measure in a $k > 2$ class problem, the sum of the paired divergences is often used as an optimization criteria [27]. Therefore, the divergence in this case is expressed

$$J_i = \sum_{l=1}^{k-1} \sum_{j=l+1}^{k} J(S_l,S_j,i), \qquad k > 2, \quad i = 1,\cdots,R \quad (17)$$

and the features are ranked one at a time by the total summed divergence.

If two features are highly correlated, they not only contain redundant information, but they make it extremely difficult to carry out statistical classification. This is particularly true of classifiers which must invert a covariance matrix. To identify the degree of correlation between features, a normalized correlation matrix C of dimension $R \times R$ was defined

$$C = [c_{ij}] = \left[\frac{\sigma_{ij}}{\sqrt{\sigma_{ii}}\sqrt{\sigma_{jj}}}\right], \qquad i,j = 1,\cdots,R. \quad (18)$$

The σ terms are the (i,j)th elements of a covariance matrix. This matrix was computed for all R features.

The final N features were chosen by first ordering the R original features by the divergence measure and accepting only features whose absolute value of c_{ij}, $i \neq j$, with all previously accepted features did not exceed 0.8. Features were added to the final set in this manner until the classification accuracy during training no longer improved.

TABLE II
SELECTED SPATIAL AND FREQUENCY DOMAIN FEATURES

Selected Spatial Measures	2 Class	$R_4(3), \bar{M}_3(1), V_4(3)\ \bar{M}_5(3), V_5(3)\ \bar{M}_5(1)$
	4 Class	$R_4(3)\ \bar{M}_5(1)\ R_5(7), V_4(3), \bar{M}_5(3)V_5(3)\ \bar{M}_1(3)$
Selected annular ring measures in line pairs/mm (lp/mm)	2 Class	.50, 4.75, .78, 2.50
	4 Class	.50, .95, 4.75, 2.50, 3.12, 7.60

The $k = 2$ class situation consisted of a normal–abnormal determination, where any film above major profusion category 0 was deemed abnormal. The $k = 4$ class problem consisted of separation of the four major small opacity profusion categories with the exception that large opacities were grouped with small opacities of major category 3. The latter compromise was necessary because of a lack of large-lesion cases in the chosen film set. In this manner, an R-dimensional feature space was reduced to an N-dimensional feature space. The selected features for the two- and four-class diagnostic classification using the digital or optical measures is shown in Table II.

It was discovered that the selected textural feature measures for normal zonal films had a variance that was small relative to the variance of the feature measures by abnormal films. This fact and the high correlation between features seemed to preclude distribution-free classifiers and led to a statistical approach. Statistical classifiers make assumptions about the underlying distributions of features. The most common assumption made about feature statistics is that they are multivariate Gaussian. The Gaussian assumption was validated via *a posteriori* classification results. Thus the discriminant function which maximizes the likelihood of a correct decision is

$$g_k(\bar{x}) = -\tfrac{1}{2}\bar{x}^t[\phi_k]^{-1}\bar{x} + \bar{x}^t[\phi_k]^{-1}\bar{\mu}_k - \tfrac{1}{2}\mu_k{}^t[\phi_k]^{-1}\bar{\mu}_k + \ln P(S_k) - \tfrac{1}{2}\ln(\det[\phi_k]), \qquad k = 1,2 \quad \text{or} \quad k = 1,\cdots,4 \quad (19)$$

where $\bar{\mu}_k$ is an N-dimensional column vector, ϕ_k is the $N \times N$ covariance matrix, and $P(S_k)$ is the *a priori* probability of a sample belonging to diagnostic class k. In a biomedical application, it is usually desired that there be a very low probability of assigning a normal classification to an abnormal subject. The discriminant function may easily be adjusted to take this into consideration by modification of $P(S_k)$.

COMPUTER CLASSIFICATION RESULTS

The computer or physician diagnostic results are all discussed in terms of confusion matrices [28], the general form of which are shown in Figs. 6 and 7. Within this context, a *true* delineation indicates that true class and

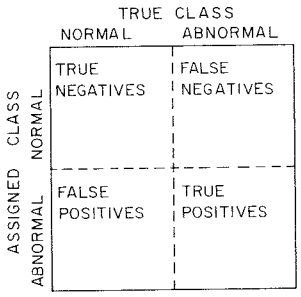

Fig. 6. Two-class confusion matrix.

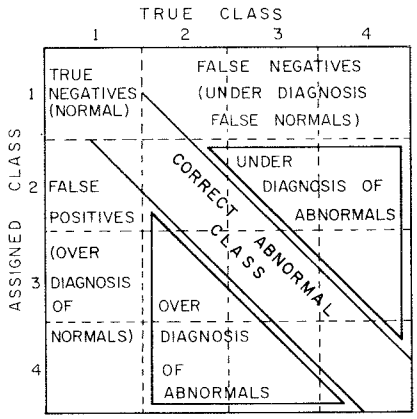

Fig. 7. Four-class confusion matrix.

assigned class were identical, whereas a *false* delineation indicates a disagreement between true and assigned classification. Within this context a true negative implies a normal that was correctly classified as such. Similarly, a false positive denotes a normal which was incorrectly classified as abnormal. A false negative implies an abnormal which was incorrectly assigned to the normal category.

The computer diagnostic testing procedure consisted of removing one sample from the data base, training on the remaining samples, and resubmitting the withdrawn sample for reclassification. This was a fair test since the classifier did not "see" the withdrawn sample until it was asked to diagnostically assign it to a class. A second more severe test was also performed in the two-class case. This test consisted of removing *one-half* of the data from each class and training on the remaining data. The removed half was then submitted to the classifier for diagnosis. This was repeated twice so that all data were classified in a test situation. The second test was not performed in the four-class diagnosis because such a procedure would be undertrained. In many respects these two testing procedures were logical extremes.

When discussing digital accuracy rates on a per-zonal-film rather than a per-interrib-space basis, the following rules of correspondence were applied. If a film contained 2 interrib spaces, the assigned diagnostic class for the film was chosen as the highest numbered of the two classes present on the film. If a film contained 3 interrib spaces, and 2 were spaces assigned to the same class, then the film was placed in that class. If all 3 spaces were classed differently, the highest numbered diagnostic class present was chosen for the film. If a film contained 4 interrib spaces, a majority of 3 in any class would assign the film to that class. If this majority did not exist, the film was assigned to the abnormal class. Therefore, an effort was made to overdiagnose the film when a consensus was not possible. There were 7 films with 2 interrib spaces, 75 with 3 spaces, and 18 with 4 spaces.

The one-at-a-time removal test procedure on digitally derived features yielded the following confusion matrix:

assigned class \ true class	1	2
1	92.6	3.6
2	7.4	96.4

The computed normal–abnormal diagnostic rate was 95.2 percent. On a per-film basis, 3 films were missed for a corresponding diagnostic rate of 96.9 percent. When the more severe second test was performed, the following confusion matrix was found:

assigned class \ true class	1	2
1	90.0	5.2
2	10.0	94.8

The normal–abnormal rate was 92.9 percent. On a per-film basis this was 96.8 percent. The 1 missed abnormal was of profusion category 1/1.

One quite obvious conclusion is that the normal–abnormal diagnostic rates were quite stable, using digitally derived textural features. The two-class classification results using transform domain features will now be presented. A corresponding classification rate will be given for the 95 films common between the two film bases. For the first optical test the results were

assigned class \ true class	1	2
1	74.4	2.9
2	25.6	97.1

The normal–abnormal rate was 90.8 percent. Of the 10 missed normal films 6 were from profusion category 0/1. The 3 missed abnormals were from profusion category 1/1. When a common film base with digitally derived features was compared, 5 out of 95 films were missed for a diagnostic normal–abnormal rate of 94.8 percent. The more severe test procedure of the second optical text yielded the following matrix:

assigned class \ true class	1	2
1	69.7	4.0
2	30.3	96.0

The normal–abnormal diagnostic rate was 88.7 percent. Of the 12 missed normals 6 were again from profusion category 0/1. The 4 missed abnormals were all of profusion category 1/1. A common film base with the digital film base once again indicated 5 misses out of 95 for a normal–abnormal rate of 94.8 percent.

Four-class digital testing results on a per-film basis yielded the following confusion matrix:

assigned class \	1	2	3	4
1	97.0	7.1	0	3.2
2	3.0	35.7	15.0	9.7
3	0	42.9	80.0	45.2
4	0	14.3	5.0	41.9

The overall four-class correct rate was 65.5 percent. The false positive rate was 3.0 percent, and the false negative rate was 3.2 percent. The normal–abnormal diagnostic rate was 96.8 percent. There was significant overdiagnosis of class 2 and underdiagnosis of class 4.

The four-class optical diagnostic testing results based on one-at-a-time removal were

assigned class \	1	2	3	4
1	74.4	15.2	3.6	2.4
2	20.5	42.4	21.4	26.8
3	0.0	9.1	32.1	26.8
4	5.1	33.3	42.9	44.0

The overall four-class correct classification was 49.6 percent. The false positive rate was 25.6 percent and the false negative rate was 6.8 percent. The normal–abnormal rate was 88 percent. When only the 95-film digital subset was considered, the overall four-class diagnostic rate was 61.0 percent with a false positive rate of 9.3 percent and a false negative rate of 1.6 percent. The normal–abnormal rate was 96.0 percent.

Physician Diagnosis

Six radiologists were requested to diagnose the identical 141 lung regions submitted for automatic analysis. Two of the six readers originally selected the films in this study. One of these two initial readers also participated in the selection of ACR standard films and is a C reader. The other initial reader is an experienced B reader who has published extensively in various aspects of chest radiographic interpretation and screens approximately 100 chest films per day for the County of Los Angeles. The other four readers consisted of one C reader and three experienced B readers. As a group these six radiologists represent over 130 man-years of radiological reading experience.

The physicians viewed entire radiographs with the 141 lung zones within which they were to make their diagnosis labeled and numbered. They were asked to disregard any information not included in the delineated zone. This procedure allowed the readers to grade individual zones within an anatomical context.

The performance of the best manual grader was

assigned class \ true class	1	2
1	94.9	1.0
2	5.1	99.0

The normal–abnormal diagnostic rate was 97.9 percent. It should be noted that this individual was one of the original film selectors. By contrast, the lowest performance of any reader was

assigned class \ true class	1	2
1	41.0	1.0
2	59.0	99.0

In this instance the normal–abnormal diagnostic rate was 83.0 percent.

When all the $6 \times 141 = 846$ physician observations were averaged, the following confusion matrix was formed:

assigned class \ true class	1	2
1	82.5	2.6
2	17.5	97.4

In this instance the normal–abnormal rate was 93.4 percent. The corresponding confusion matrix in the four-class physician diagnosis was

assigned class \	1	2	3	4
1	82.5	7.1	0.6	0.4
2	17.1	57.1	11.9	4.5
3	0.4	24.7	53.0	22.0
4	0	11.1	34.5	73.2

The overall correct classification was 68.0 percent. When the two original film selectors were excluded, this rate dropped to 67.0 percent. The false positive rate also rose to 24.5 percent with no significant change in false negative rate. It was also discovered that the two original readers were in agreement with *themselves* 96 percent of time in a normal–abnormal sense and 77 percent of the time in a four-class sense. This was a useful measure of intrareader variation.

In summary, the false positive rates for the six readers ranged from 59.0 to 2.9 percent with an average of 17.5 percent. Correspondingly, the false negative rate ranged from 1.0 to 6.9 percent with an average of 2.6 percent. The averaged physician rates showed no significant changes when computed on the basis of the 95 films submitted for digital analysis.

Conclusions

This study was undertaken as a short-term project to determine the feasibility of the possible automated mass diagnostic screening of pneumoconiosis radiographs. Two distinct textural feature extraction methods involving digital and coherent optical approaches were undertaken. The performance of the two automated diagnostic systems was described in detail and tables presented. Analogous results were presented for diagnoses obtained from experienced radiologists asked to analyze the same films given to the automated systems.

As this was a feasibility study, the available data base was necessarily limited. As the data base is expanded and statistics of the measured features become better known, it may be conjectured that performance will tend to improve. Nevertheless, performance was most encouraging.

Automated normal–abnormal classification accuracy on a testing basis was no lower than 88.0 percent. The comparable percentages for the physicians ranged from 83.0 percent to 97.9 percent. Likewise, the false normal computer testing rate never exceeded 4 percent, with a physician range of 1 to 7 percent. The false positive rate never exceeded 30 percent, as compared to a physician range of 5.2 to 59.0 percent. While these initial results are encouraging, much further research and development needs to be accomplished before an automated film reader for this application will become a reality. It is felt that if a device is to perform this task it must be certified to read in a manner and with a set of films *identical* to a set used to certify future B readers. Such an acceptance test does not exist at present but there is reason to suspect that one will soon be devised.

Each of the automated approaches discussed had its own special advantages. The hybrid optical–digital system is particularly fast. As a result, the per-film expense of diagnostic screening should be relatively low. The purely digital system, while considerably slower, is more flexible and able to compute and analyze data in addition to textural measures over the lung field. In particular, such features as anatomical structure may be dealt with in a unified manner. There is also evidence that digital measures may yield a better severity diagnosis.

This suggests a two-stage approach. Initial screening could be done optically. A second step would analyze the initially selected abnormal zones digitally in greater detail. Final determination of normal–abnormal classification would then be made, and an estimate of the severity of the disease given in abnormal cases.

Acknowledgment

The authors wish to acknowledge the cooperation of Mr. N. Jensen and the staff of Recognition Systems, Inc., for the use of the ROSA-3 device and Dr. G. Jacobson, Chairman of the Department of Radiology, University of Southern California, and his staff for their assistance and advice. We would also like to acknowledge Profs. G. Bekey and W. Pratt of the Biomedical Engineering and Image Processing Institutes for their cooperation and facilities, as well as Prof. H. C. Andrews for his advice on pattern recognition techniques.

References

[1] H. Bohlig *et al.*, "UICC/Cincinnati classification of the radiographic appearances of the pneumoconioses," *Chest*, vol. 58, pp. 57–67, July 1970.
[2] M. M. Key and L. E. Kerr, Eds., *Pulmonary Reactions to Coal Dust*. New York: Academic Press, 1971.
[3] G. Jacobson and W. S. Lainhart, Eds., "ILO U/C 1971 International classification of radiographs and pneumoconioses," *Med. Radiogr. Photogr.*, vol. 48, no. 3, pp. 66–113, 1972.
[4] F. D. K. Liddell, "Validation of classification of pneumoconiosis," presented at the Int. Symp. Coal Worker's Pneumoconiosis, New York Acad. Sci., Sept. 1971.
[5] C. E. Rossiter, D. Rivers, I. Bergnan, C. Casswell, and G. Nagelschmidt, "Dust content, radiology and pathology in simple pneumoconiosis of coal workers," in *Inhaled Particles and Vapors II*, C. N. Davis, Ed. New York: Pergamon, 1967, sect. 5, pp. 419–437.
[6] R. T. Moore, "NIOSH report," presented at the Meeting on Radiographic Description of the Pneumoconioses, Cleveland, Ohio, Sept. 1972.
[7] L. H. Garland, "Studies on the accuracy of diagnostic procedures," *Amer. J. Roentgenol. Radium Ther. Nucl. Med.*, vol. 82, pp. 25–88, July 1959.
[8] J. Yerushalmy, "The statistical assessment of the variability in observer perception and description of roentgenographic pulmonary shadows," *Radiol. Clin. N. Amer.*, vol. VII, pp. 381–342, Dec. 1969.
[9] R. B. Reger and W. K. C. Morgan, "On the factors influencing consistency in radiologic diagnosis of pneumoconiosis," *Amer. Rev. Resp. Dis.*, vol. 102, pp. 905–915, 1970.
[10] R. B. Reger, J. Smith, A. Kibelstis, and W. K. C. Morgan, "The effect of film quality and other factors on roentgenographic categorization of coal workers pneumoconiosis," *Amer. J. Roetgenol. Radium Ther. Nucl. Med.*, vol. 115, pp. 462–472, July 1972.
[11] G. Jacobson, H. Bohlig, and R. Kiviluoto, "Essentials of chest radiography," *Radiology*, vol. 95, pp. 445–450, May 1970.
[12] R. P. Kruger, "Computer processing of radiographic images," Ph.D. dissertation, Univ. Missouri, Columbia, Mo., June 1971.
[13] R. P. Kruger, J. R. Townes, D. L. Hall, S. J. Dwyer, III, and G. S. Lodwick, "Automated radiographic diagnosis via feature extraction and classification of cardiac size and shape descriptors," *IEEE Trans. Bio-Med. Eng.*, vol. BME-19, pp. 174–186, May 1972.
[14] A. Rosenfeld and E. Troy, "Visual texture analysis," in *Proc. UMR–Kelly Communications Conf.*, Rolla, Mo., 1970.
[15] R. M. Pickett, "Visual analysis of texture in the detection and recognition of objects," in *Picture Processing and Psychopictorics*, B. C. Lipkins and A. Rosenfeld, Eds. New York: Academic Press, 1970, pp. 289–308.
[16] J. J. Gibson, "The perception of visual surfaces," *Amer. J. Psychol.*, vol. 63, pp. 367–384, 1950.
[17] J. K. Hawkin, "Textural properties for pattern recognition," in *Picture Processing and Psychopictorics*, B. C. Lipkins and A. Rosenfeld, Eds. New York: Academic Press, 1970, pp. 347–370.
[18] H. C. Andrews, A. G. Tescher, and R. P. Kruger, "Image processing by digital computer," *IEEE Spectrum*, vol. 9, pp. 20–32, July 1972.
[19] A. Rosenfeld and E. Troy, "Visual texture analysis," Univ. Maryland, College Park, Tech. Rep. 70–116, June 1970.
[20] R. Haralick and D. Anderson, "Textural–tone study with applications to digitized memory," Univ. Kansas, Lawrence, Tech. Rep. 182-2, Nov. 1971.
[21] D. A. Ausherman, "Texture discrimination within digital imagery," Ph.D. dissertation, Univ. Missouri, Columbia, Dec. 1972.
[22] G. G. Lendaris and G. L. Stanley, "Diffraction pattern sampling for automatic pattern recognition," *Proc. IEEE*, vol. 58, pp. 198–216, Feb. 1970.
[23] E. L. Hall, R. P. Kruger, S. J. Dwyer, III, D. L. Hall, R. W. McLaren, and G. S. Lodwick, "A survey of preprocessing and feature extraction techniques for radiographic images," *IEEE Trans. Comput.*, vol. C-20, pp. 1032–1044, Sept. 1971.
[24] R. N. Sutton and E. L. Hall, "Texture measures for automatic classification of pulmonary disease," *IEEE Trans. Comput.*, vol. C-21, pp. 667–676, July 1972.
[25] R. H. Morgan, L. M. Bates, U. V. Gopalarao, and A. Marinaro, "The frequency response characteristics of X-ray films and screens," *Amer. J. Roentgenol. Radium Ther. Nucl. Med.*, vol. 92, pp. 426–440, Aug. 1964.
[26] K. Rossmann, "Measurement of the modulation transfer function of radiographic systems containing fluorescent screens," *Phys. Med. Biol.*, vol. 9, no. 4, pp. 551–557, 1964.
[27] H. C. Andrews, *Introduction to Mathematical Techniques in Pattern Recognition*. New York: Wiley, 1972.
[28] L. B. Lusted, "Perception of the roentgen image: applications of signal detectability theory," *Radiol. Clin. N. Amer.*, vol. 7, pp. 435–445, Dec. 1969.

Reprinted by permission from IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS
Vol. SMC-4, No. 1, January 1974, pp. 40-49
Copyright 1974, by the Institute of Electrical and Electronics Engineers, Inc.
PRINTED IN THE U.S.A.