Applications of fuzzy languages and pictorial databases to decision support systems design

by EDWARD T. LEE
Memphis State University
Memphis, Tennessee

ABSTRACT

The pioneering work of D. T. Lee in developing an approach with major emphasis on database development has had a profound influence on the recent development of decision support systems, as well as on office information systems, database systems, and database machines. This database development approach is a new and powerful approach to decision support systems design methodologies.

In this paper the concepts of fuzzy languages and pictorial databases are applied to decision support systems design methodologies. First, fuzzy languages, fuzzy grammars, the classification of fuzzy grammars, derivation chain, degree of acceptance, and equivalence are defined. Operations like intersection, concatenation, Kleene closure, complement, and cardinality are also defined. Second, algebraic representation of the production system is presented and illustrated by examples. The difference between null string and the empty set of string is illustrated. Third, decision support systems involving geometric figures, chromosome images or leukocyte images are presented as illustrative examples.

In similarity retrieval from a pictorial database, very often it is desired to find pictures (or feature vectors, histograms, etc.) that are most similar to or most dissimilar to a test picture (or feature vector). Using similarity measures, one can not only store similar pictures logically or physically close to each other to improve retrieval or updating efficiency, one can also use such similarity measures to answer fuzzy queries involving nonexact retrieval conditions.

The applications of fuzzy languages and pictorial databases to decision support systems design methodologies offer what appears to be a fertile field for further study. The underlying ideas are interesting and easy for practical application. The results have useful applications in decision support systems, pattern recognition, pictorial information systems, and artificial intelligence.
Applications of Fuzzy Languages and Pictorial Databases to DSS Design

A. INTRODUCTION

D. T. Lee's pioneering work in developing an approach that emphasizes database development has profoundly influenced the recent development of decision support systems, as well as office information systems, database systems, and database machines. This database development approach is a new and powerful approach to designing decision support systems.

During the past several years, fuzzy languages and similarity retrieval techniques have attracted growing attention as promising avenues of approach to problems in decision support systems design methodologies.

B. FUZZY LANGUAGES

In the theory of formal languages, a language $L$ is defined as a subset of $V^*_T$—as the set of strings over a finite alphabet $V_T$. Consequently, if $x$ is a string in $V^*_T$ then either $x$ is a member of $L$ or $x$ is not a member of $L$. In contrast, in the case of a natural language, a sentence may be partially in a language, in the sense that it may be partially grammatically correct (or partially meaningful). Indeed, it may be argued that it is the complete precision of formal languages that sets them so sharply apart from natural languages.

The gap between formal and natural languages can be narrowed by introducing the concept of fuzzy language, that is, a fuzzy subset of $V^*_T$. In such a language, each string $x \in V^*_T$ is assigned a grade of membership, $\mu_L(x)$, in $L$, which for simplicity may be taken to be a number in the interval $[0, 1]$. When $L$ is a natural language, $\mu_L(x)$ may be interpreted as an index of the grammatical correctness of $x$.

As was pointed out by E. T. Lee and Zadeh, much of the existing theory of formal languages can be extended quite readily to fuzzy languages. This has been done for the notions of grammar, Kleene closure, recursiveness, context-sensitive languages, context-free languages, regular languages, Chomsky normal form, Greibach normal form, and so on. In the present paper, we shall focus our attention on the applications of fuzzy languages to decision support systems design methodologies. As in Lee and Zadeh, our notation, terminology and constructions for fuzzy languages will parallel closely those of Hopcroft and Ullman for nonfuzzy formal languages.

We begin with a brief recapitulation of some of Lee and Zadeh's definitions.

Let $V_T$ be a set of terminals, with $V^*_T$ denoting the set of strings over $V_T$. A fuzzy language $L$ is a fuzzy subset of $V^*_T$. As such, it may be characterized by a membership function $\mu_L : V^*_T \rightarrow [0, 1]$ that associates with each string $x$ in $V^*_T$ its grade of membership, $\mu_L(x)$, in $L$.

If $L_1$ and $L_2$ are two fuzzy languages, then their union is a fuzzy language denoted by $L_1 \cup L_2$ and defined by

$$\mu_{L_1 \cup L_2}(x) = \max\{\mu_{L_1}(x), \mu_{L_2}(x)\}, \quad x \in V^*_T,$$

or more compactly

$$\mu_{L_1 \cup L_2} = \mu_{L_1} \vee \mu_{L_2},$$

where $\vee$ stands for Max (on infix form). To simplify the notation, we shall write throughout $\vee$ for Max, $\wedge$ for Min and will omit the arguments of membership functions when an equality or inequality that these functions satisfy holds for all strings in $V^*_T$.

The intersection of $L_1$ and $L_2$ is denoted by $L_1 \cap L_2$ and is defined by

$$\mu_{L_1 \cap L_2} = \mu_{L_1} \wedge \mu_{L_2}.$$  

The concatenation of $L_1$ and $L_2$ is denoted by $L_1L_2$ and is defined by

$$\mu_{L_1L_2}(x) = \bigvee_u (\mu_{L_1}(u) \wedge \mu_{L_2}(x-u)), \quad x \in V^*_T,$$

where $u$ and $x-u$ denote, respectively, a prefix and the corresponding suffix of $x$.

The Kleene closure of a language $L$ is defined by

$$L^* = \xi + L + LL + LLL + \ldots$$

where $\xi$ denotes the null string.

The complement of a language $L$ is denoted by $\overline{L}$ and is defined by

$$\mu_{\overline{L}} = 1 - \mu_L.$$  

The cardinality of a language $L$ is denoted by

$$|L| = \sum_i \mu_L(x_i).$$

A fuzzy grammar is a quadruple $G = (V_N, V_T, P, S)$ where $V_N$ is a set of non-terminals (i.e., labels for certain fuzzy subsets of $L$), $P$ is a set of fuzzy productions and $S \in V_N$. A generic fuzzy production has the form

$$\alpha \xrightarrow{\rho} \beta,$$

where $\alpha, \beta \in (V_T \cup V_N)^*$ and $0 < \rho \leq 1$.  

From the collection of the Computer History Museum (www.computerhistory.org)
If \( \alpha_1, \ldots, \alpha_m \) are strings in \( (V_T \cup V_N)^* \) and

\[
\begin{align*}
\alpha_1 & \xrightarrow{p_1} \alpha_2, \\
\alpha_{m-1} & \xrightarrow{p_m} \alpha_m,
\end{align*}
\]

then \( \alpha_m \) is derivable from \( \alpha_1 \) in grammar \( G \). This is expressed as

\[
\alpha_1 \Rightarrow \alpha_m
\]

or simply \( \alpha_i \Rightarrow \alpha_m \). The expression

\[
\alpha_1 \xrightarrow{P} \alpha_2 \ldots \xrightarrow{p_m} \alpha_m
\]

is called a derivation chain from \( \alpha_1 \) to \( \alpha_m \). The strength of such a chain is defined to be \( p_2 \wedge \ldots \wedge p_m \), which may be interpreted as the strength of its weakest link.

A fuzzy grammar \( G \) generates a fuzzy language \( L(G) \) in the following manner. A string of terminals \( x \) is in \( L(G) \) if and only if \( x \) is derivable from \( S \). The grade of membership of \( x \) in \( L(G) \) is given by

\[
\mu_G(x) = \sup \{ p_1 \wedge \ldots \wedge p_{m+1} \}
\]

where \( S \xrightarrow{P} \alpha_1 \ldots \xrightarrow{p_m} \alpha_m \xrightarrow{p_{m+1}} x \) is a derivation chain from \( S \) to \( x \) and the supremum is taken over all derivation chains from \( S \xrightarrow{P} x \). In words,

\[
\mu_G(x) = \text{strength of the strongest derivation chain from } S \text{ to } x.
\]

Thus, (10) defines \( L(G) \) as a fuzzy set in \( (V_T \cup V_N)^* \). If \( L(G_1) = L(G_2) \) in the sense of equality of fuzzy sets, then the grammars \( G_1 \) and \( G_2 \) are said to be equivalent.

This concludes the recapitulation of some of the basic concepts defined by Lee and Zadeh.\(^{12}\)

C. ALGEBRAIC REPRESENTATION OF THE PRODUCTION SYSTEM

For many purposes it is convenient to replace \( P \) by a set of algebraic equations involving the operations of concatenation and addition (union) of fuzzy sets of strings. Such equations constitute an extension to fuzzy sets of strings of the approach used by Rosenkrantz to derive the Greibach normal form.\(^{13}\)

The replacement is effected as follows. If \( P \) contains productions of the form

\[
\begin{align*}
\alpha & \xrightarrow{p_1} \beta, \quad \alpha, \beta \in (V_T \cup V_N)^* \\
\alpha & \xrightarrow{p_2} \gamma
\end{align*}
\]

then

\[
\alpha = p_1 \beta + p_2 \gamma,
\]

where \( + \) denotes the union of the fuzzy sets of strings represented by \( p_1 \beta \) and \( p_2 \gamma \).

Example 1. Written in algebraic form, the production system of

\[
\begin{align*}
S & \xrightarrow{0.5} AB \\
S & \xrightarrow{0.8} A \xrightarrow{0.6} b \\
S & \xrightarrow{0.8} B \xrightarrow{0.4} A \\
S & \xrightarrow{0.8} B \xrightarrow{0.2} a \\
S & \xrightarrow{0.5} a \\
AB & \xrightarrow{0.4} BA
\end{align*}
\]

reads

\[
\begin{align*}
S &= 0.5AB + 0.8A + 0.8B \\
A &= 0.5a + 0.6b \\
B &= 0.4A + 0.2a \\
AB &= 0.4BA
\end{align*}
\]

As in the case of nonfuzzy languages, it is convenient to classify the grammars of fuzzy languages into four principal categories, which follow, in order of decreasing generality.

C.1 Type 0 Grammars

In this case, productions are of the general form

\[
\alpha \xrightarrow{p} \beta,
\]

where \( \alpha \) and \( \beta \) are strings in \( (V_T + V_N)^* \), with \( \alpha \neq \xi \).

C.2 Type 1 Grammars (Context Sensitive)

Here the productions are of the form

\[
\alpha \xrightarrow{p} \beta,
\]

where \( \alpha \) and \( \beta \) are strings in \( (V_T + V_N)^* \), with \( \alpha \neq \xi \) and \( |\beta| \geq |\alpha| \), that is, the length of the right-hand side (the consequent) must be at least as great as the length of the left-hand side (the antecedent).

C.3 Type 2 Grammars (Context Free)

Here the allowable productions are of the form

\[
A \xrightarrow{p} \alpha
\]

where \( A \in V_N \), \( \alpha \in (V_T + V_N)^* \), and \( S \xrightarrow{p} \xi \) is allowed. Thus, in the case of a context-free grammar, \( A \) can be replaced by \( \alpha \) regardless of the context in which \( A \) occurs.

C.4 Type 3 Grammars (Regular)

In this case the allowable productions are of the form

\[
\begin{align*}
A & \xrightarrow{p} aB \\
A & \xrightarrow{p} a \\
S & \xrightarrow{p} \xi
\end{align*}
\]
where \( A, B \in V_N \) and \( a \in V_T \).

In solving a system of algebraic equations representing the production system of a fuzzy grammar, one frequently encounters linear equations of the form

\[ X_i = P_i X_i + Q_i, \quad (12) \]

in which \( X_i, P_i, \) and \( Z_i \) are fuzzy sets of strings over a finite alphabet, and + and the product denote the union and the concatenation, respectively.

Let \( L_\emptyset = \{ \emptyset \} = \text{null string}; \)

\[ LL_\emptyset = L_\emptyset L = L. \quad (13) \]

\( L_\emptyset \) plays the role of the unit element for concatenation.

Let \( \emptyset = \text{the empty set of strings}, \) then

\[ A \emptyset = \emptyset A = \emptyset \quad (14) \]

since there are no elements in \( \emptyset. \)

**Proposition 1.** If \( P_i \) does not contain the null string, then (12) has a unique solution for \( X_i \), which is given by

\[ X_i = P_i^* Q_i, \quad (15) \]

where \( P_i^* \) is the Kleene closure of \( P_i \) (in the sense of (5)).

The algebraic notation that was described earlier is particularly useful in the case of context-free grammars. Thus if the nonterminals in \( V_N \) are denoted by \( X_1, \ldots, X_n \), and \( X = (X_1, \ldots, X_n) \), with \( X_1 = S \), then the production system \( P \) can be put into the form

\[ X = f(X) \quad (16) \]

where \( f \) is an \( n \)-vector whose components are multinomials in the \( X_i, i = 1, \ldots, n \). In this way, the determination of the fuzzy set of strings generated by the grammar reduces to finding a fixed point of the function \( f \). In this connection, it can readily be shown that if we set \( X^0 = \emptyset = \text{empty set} \) and form the iterates

\[ X^{k+1} = f(X^k), \quad X^0 = \emptyset, \quad k = 1, 2, 3, \ldots, \quad (17) \]

then, for each \( k \), \( X^k \) is a fuzzy subset of the solution of (16).

We are now ready to turn our attention to the applications of fuzzy languages to decision support systems design methodologies. Algorithms for finding Chomsky and Greibach normal forms for a fuzzy context-free grammar using an algebraic approach have been found. 17

**D. DECISION SUPPORT SYSTEMS INVOLVING GEOMETRIC FIGURES**

Pictures, images, and figures play basic and important roles in decision support systems.
Lemma 1. Given a triangle $\triangle ABC$ with angles $A$, $B$, $C$, if we assume that $A \geq B \geq C$, then

$$
\mu_{f}(\triangle ABC) = 1 - \frac{1}{60} \min \{ |A - B|, |B - C|, |C - A| \},
$$

$$
\mu_{E}(\triangle ABC) = 1 - \frac{1}{180} \max \{ |A - B|, |B - C|, |C - A| \},
$$

and

$$
\mu_{R}(\triangle ABC) = 1 - \frac{1}{90} \min \{ |A - 90^\circ|, |B - 90^\circ|, |C - 90^\circ| \},
$$

respectively.

A quantitative measure of the similarity of $\triangle ABC$ to isosceles right triangles may be defined as

$$
\mu_{IR} = \min \{ \mu_{f}, \mu_{R} \}
$$

or

$$
\mu_{IR} = \mu_{f} \cdot \mu_{R}.
$$

Since both $\mu_{f}$ and $\mu_{R}$ are in the range 0 to 1, $\mu_{IR} \geq \mu_{IR}$. Depending on its prospective application, $\mu_{IR}$ may be substituted for $\mu_{IR}$ or vice-versa.

A quantitative measure of the similarity of $\triangle ABC$ to scalene triangles may be defined as:

$$
\mu_{SC} = 1 - \max \{ \mu_{f}, \mu_{R}, \mu_{E} \}.
$$

For a triangle $\triangle ABC$ with $A \geq B \geq C$, we shall use the following vector representation for convenience:

$$
\triangle ABC = (A, B, C).
$$

Lemma 2. Given a triangle $\triangle ABC$, the set

$$
\{ 0, \mu_{IR}(\triangle ABC), \mu_{IR}(\triangle ABC), \mu_{f}(\triangle ABC), \mu_{R}(\triangle ABC), 1 \}
$$

with max and min as the two binary operations forms a distributive but not complemented lattice.

Example 2. Table I gives the $\mu_{f}$, $\mu_{E}$, $\mu_{R}$, $\mu_{IR}$, and $\mu_{SC}$ of 12 triangles.

In the paper by Zadeh,19 "not," "and," and "or" are interpreted as "the operation of complementation (or, equivalently, negation)," "the operation of intersection," and "the operation of union," respectively. In another paper by Zadeh,20 such linguistic hedges as "very," or "more or less," were viewed as operators that act on the fuzzy sets representing the meaning of their operands. More specifically, "very" is interpreted as "the operation of concentration," which has the effect of squaring the membership function; "more or less" is interpreted as "the operation of dilation," which has the effect of taking the square root of the membership function. Composite fuzzy queries can be answered by using linguistic hedges and quantitative fuzzy semantics.

Example 3. The grade of membership of the triangle (90°, 75°, 15°) with respect to the class "very similar to isosceles triangles" is 9/16; its grade of membership in the class "more or less similar to isosceles triangles" is $\sqrt{3}/2$.

Example 4. Assuming the 12 triangles in Table I as the database, the composite fuzzy query, "retrieve the triangles that are very similar to equilateral triangles and more or less similar to right triangles" may be answered by computing

**TABLE I—Similarity measures for 12 triangles, $\triangle ABC$**

<table>
<thead>
<tr>
<th>$\triangle ABC$</th>
<th>$\mu_{f}$</th>
<th>$\mu_{E}$</th>
<th>$\mu_{R}$</th>
<th>$\mu_{IR}$</th>
<th>$\mu_{SC}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(90°, 70°, 20°)</td>
<td>$\frac{3}{4}$</td>
<td>$\frac{4}{5}$</td>
<td>$\frac{3}{4}$</td>
<td>$\frac{1}{3}$</td>
<td>$\frac{1}{2}$</td>
</tr>
<tr>
<td>(90°, 60°, 30°)</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{2}$</td>
</tr>
<tr>
<td>(120°, 60°, 0°)</td>
<td>0</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{2}$</td>
<td>0</td>
<td>$\frac{1}{2}$</td>
</tr>
<tr>
<td>(60°, 60°, 60°)</td>
<td>1</td>
<td>1</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{2}$</td>
</tr>
<tr>
<td>(90°, 45°, 45°)</td>
<td>1</td>
<td>$\frac{1}{2}$</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>(180°, 0°, 0°)</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(75°, 60°, 45°)</td>
<td>$\frac{3}{4}$</td>
<td>$\frac{3}{4}$</td>
<td>$\frac{3}{4}$</td>
<td>$\frac{3}{4}$</td>
<td>$\frac{3}{4}$</td>
</tr>
<tr>
<td>(75°, 75°, 30°)</td>
<td>1</td>
<td>$\frac{3}{4}$</td>
<td>$\frac{3}{4}$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(90°, 75°, 15°)</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{2}$</td>
</tr>
<tr>
<td>(120°, 30°, 30°)</td>
<td>1</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{2}$</td>
</tr>
<tr>
<td>(120°, 45°, 15°)</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{2}$</td>
</tr>
<tr>
<td>(150°, 15°, 15°)</td>
<td>1</td>
<td>$\frac{3}{4}$</td>
<td>$\frac{3}{4}$</td>
<td>$\frac{3}{4}$</td>
<td>$\frac{3}{4}$</td>
</tr>
</tbody>
</table>

From the collection of the Computer History Museum ([www.computerhistory.org](http://www.computerhistory.org))
From the collection of the Computer History Museum (www.computerhistory.org)

TABLE II—Similarity measures with linguistic hedges for the 12 triangles of Table I: "Very similar to an equilateral triangle," "more or less similar to a right triangle," and "both very similar to an equilateral triangle and more or less similar to a right triangle."

<table>
<thead>
<tr>
<th>$\triangle ABC$</th>
<th>$\mu_x^2$</th>
<th>$\mu_A^2$</th>
<th>$\mu_x^2 \land \mu_A^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. (90°, 70°, 20°)</td>
<td>0.37</td>
<td>0.37</td>
<td></td>
</tr>
<tr>
<td>2. (90°, 60°, 30°)</td>
<td>0.44</td>
<td>0.44</td>
<td></td>
</tr>
<tr>
<td>3. (120°, 60°, 0°)</td>
<td>0.11</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td>4. (60°, 60°, 60°)</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>5. (90°, 45°, 45°)</td>
<td>0.56</td>
<td>0.56</td>
<td></td>
</tr>
<tr>
<td>6. (180°, 0°, 0°)</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>7. (75°, 60°, 45°)</td>
<td>0.69</td>
<td>0.91</td>
<td>0.69</td>
</tr>
<tr>
<td>8. (75°, 75°, 30°)</td>
<td>0.56</td>
<td>0.91</td>
<td>0.56</td>
</tr>
<tr>
<td>9. (90°, 75°, 15°)</td>
<td>0.34</td>
<td>1</td>
<td>0.34</td>
</tr>
<tr>
<td>10. (120°, 30°, 30°)</td>
<td>0.25</td>
<td>0.82</td>
<td>0.25</td>
</tr>
<tr>
<td>11. (150°, 45°, 15°)</td>
<td>0.18</td>
<td>0.82</td>
<td>0.18</td>
</tr>
<tr>
<td>12. (150°, 15°, 15°)</td>
<td>0.06</td>
<td>0.53</td>
<td>0.06</td>
</tr>
</tbody>
</table>

1. $\mu_x^2$, the membership function, very similar to equilateral triangles
2. $\mu_A^2$, the membership function, more or less similar to right triangles
3. $\min[\mu_x^2, \mu_A^2]$, denoted by $\mu_x^2 \land \mu_A^2$, the membership function very similar to equilateral triangles and more or less similar to right triangles

These are given in Table II.

If we set the threshold to be 0.6, then the answer to the query is triangles (60°, 60°, 60°) and (75°, 60°, 45°). Abandon of a threshold is interpreted as a threshold of 0; any element with a grade of membership greater than 0 will be part of the answer; then the answer to our fuzzy query is the fuzzy set $\{(\triangle ABC), \mu_x^2 \land \mu_A^2(\triangle ABC)\}$, where $\mu_x^2 \land \mu_A^2$ is the membership function.

Zadeh discusses the classification of linguistic hedges and the operations of contrast intensification, fuzzification, and accentuation.20

Elsewhere, for $\triangle x$ with angles $A \geq B \geq C$ and $\triangle y$ with angles $A' \geq B' \geq C'$, we have set the similarity between $\triangle x$ and $\triangle y$ equal to

$$\mu_x(\triangle x, \triangle y) = 1 - \frac{1}{240^\circ} (|A - A'| + |B - B'| + |C - C'|)$$

Example 5. Given a triangle $\triangle y$ with angles $A' \geq B' \geq C'$ and a tolerance $\epsilon$ with $0 \leq \epsilon \leq 1$, the fuzzy query "retrieve all the triangles that are similar to $\triangle y$ within a tolerance of $\epsilon$" may be carried out as follows:

Let $X$ be the set of triangles $\triangle x$ in the database with angles $A \geq B \geq C$. Since angles $A$, $B$, $C$ are ordered from greatest to least, the ranges for the angles must be

$$60^\circ \leq A \leq 180^\circ$$
$$0^\circ \leq B \leq 90^\circ$$
$$0^\circ \leq C \leq 60^\circ$$

Thus, we obtain

$$60^\circ \leq A \leq 148^\circ$$

and

$$\mu_A(\triangle x', \triangle y) \equiv 0.8$$

If $\epsilon = 0.10$ then the set of possible candidates becomes those triangles where $60^\circ \leq A \leq 124^\circ$.

A dissimilarity measure of $\triangle x$ and $\triangle y$ may be defined as the complement of the similarity of $\triangle x$ and $\triangle y$:

$$\mu_{DIS}(\triangle x, \triangle y) = 1 - \mu_A(\triangle x, \triangle y)$$

Example 7. Given two triangles $\triangle x$ and $\triangle y$, the fuzzy query "retrieve all the triangles that are more or less dissimilar to $\triangle x$ or very very similar to $\triangle y", may be answered by computing $\mu^{\triangle y}_{DIS}(\triangle x, \triangle y)`$ for all triangles $\triangle x$ in the database, where $\lor$ is an infix operator for max.

Therefore, other composite fuzzy queries involving chromosomes or leukocytes can also be answered in the same manner. Examples of composite fuzzy queries involving chromosomes are "retrieve the chromosomes that are very similar to median chromosomes, but not similar to a given chromosome," and "retrieve the chromosomes that are more or less similar to median chromosomes and very very similar to a given chromosome." Examples of composite fuzzy queries involving leukocytes are "retrieve leukocytes with deeply indented nucleus or round nucleus" and "retrieve leukocytes with elongated nucleus or slightly indented nucleus."

In what follows, propositions are presented for an effective way of organizing a shape-oriented triangle database.
Proposition 2. For shape-oriented storage of triangles, it is advantageous to store the angles of a triangle in decreasing order of magnitude. This representation may be viewed as a normal form for shape-oriented triangle representation.

Proposition 3. For shape-oriented storage of triangles, if we logically order all the triangles individually and independently according to the magnitude of the angles \( A, B, \) and \( C \), then we can reduce retrieval time for answering queries.

Proposition 4. For shape-oriented storage of triangles, we can use associative memory with match-on-between-limits and find-match-count operations \( Z_1 \) or other hardware searching facilities such as the Symbol-2R computer \( Z_2 \) to reduce retrieval time.

E. DECISION SUPPORT SYSTEM INVOLVING CHROMOSOME IMAGES

A preliminary study of applying shape-oriented similarity measures defined over a pair of chromosome images to the classification problem has been presented. In this paper, the classification problem is studied through the use of shape-oriented measures of the similarity of a given chromosome image to symmetrical chromosomes, median chromosomes, submedian chromosomes, and acrocentric chromosomes.

The best-fit skeletal "length and angle only" transformation with angles \( A_i \) and sides \( a_i \) is shown in Figure 3. After connecting the tips of arms no. 1 and 2, and the tips of arms no. 3 and 4, as indicated by dotted lines, the best-fit skeletal "length and angle only" transformation becomes a hexagon.

Figure 3—Chromosome images: (a) median, (b) submedian, (c) acrocentric, (d) best-fit skeletal "length and angle only" transformation

E.2 Symmetry of Chromosome Images

The preparation of chromosome images and the definition of metaphase chromosome images is discussed by Widrow. At metaphase each chromosome has a twin, normally identical counterpart.

Definition 1. A chromosome image with angles \( A_i \) and sides \( a_i \) is a symmetrical chromosome image if and only if \( A_{2i-1} = A_{2i} \) for \( 1 \leq i \leq 4 \), \( a_1 = a_2 \) and \( a_3 = a_4 \).

A shape-oriented quantitative measure of the similarity of a given chromosome image \( A \) to all symmetrical chromosome images may be defined as

\[
\mu_s(A) = 1 - \rho_s \sum_{i=1}^{4} |A_{2i-1} - A_{2i}|
\]

where \( \rho_s \) is a normalization constant to be determined.

Lemma 3. A chromosome image is a symmetrical chromosome image if and only if \( \mu_s = 1 \).

In what follows, we assume that

\[ A_i \leq 180^\circ \text{ for } 1 \leq i \leq 8. \]

The angles of chromosome image \( A \) are represented in a vector form as

\[ A = (A_1, A_2, A_3, A_4, A_5, A_6, A_7, A_8). \]

E.2 Determination of the Normalization Constant \( \rho_s \)

The purpose of \( \rho_s \) is to normalize the value of \( \mu_s \) between 0 and 1. In order to determine the value of \( \rho_s \), we must first determine \( \text{max}_{A} \{ \sum_{i=1}^{4} |A_{2i-1} - A_{2i}| \} \). After we find this value, we assume \( \mu_s \) to be equal to zero at this extreme case, so that we can determine the value of \( \rho_s \).

Lemma 4. For any chromosome image \( A \) with angles \( A_i \),

\[ \text{max}_{A} \{ \sum_{i=1}^{4} |A_{2i-1} - A_{2i}| \} = 720^\circ. \]

Definition 2. A chromosome image is a most unsymmetrical chromosome image if and only if \( \mu_s = 0 \).

Theorem 1

\[ \rho_s = \frac{1}{720^\circ}. \]

Proof. Assuming that the symmetry measure \( \mu_s \) of a most unsymmetrical chromosome image is equal to zero, then

\[ \mu_s = 0 = 1 - \rho_s \cdot 720^\circ \text{ and } \rho_s = \frac{1}{720^\circ}. \]

Corollary 1. Given a chromosome image \( A \) with angles \( A_i \),

\[ \mu_s(A) = 1 - \frac{1}{720^\circ} \sum_{i=1}^{4} |A_{2i-1} - A_{2i}|. \]
E.3 Most Unsymmetrical Chromosome Images

Shape-oriented quantitative measures of the dissimilarities of a given triangle to isosceles triangles, right triangles, and equilateral triangles have been defined. According to the measures defined, it was also proved that the triangle most dissimilar to an equilateral triangle is \((120°, 60°, 0°)\); the triangle most dissimilar to an isosceles triangle is \((180°, 0°, 0°)\); and the triangle most dissimilar to a right triangle is \((180°, 0°, 0°)\). It is of interest to ask whether the most unsymmetrical chromosome image is unique or not. If not, then what are all the most unsymmetrical chromosome images? In order to answer this question in a concise form, we need to define an equivalence relation among chromosome images. We first define the following four permutations of the angles

\[
P_0(A) = A
\]

\[
P_T(A) = (A_2, A_1, A_4, A_3, A_6, A_5, A_8, A_7)
\]

\[
P_R(A) = (A_4, A_3, A_2, A_1, A_6, A_5, A_8, A_7)
\]

\[
P_{TR}(A) = (A_3, A_4, A_1, A_2, A_6, A_5, A_8, A_7)
\]

\[
P_0\) is the identical permutation. \(P_T\) is the permutation that interchanges the right-hand-side angles and the left-hand-side angles. \(P_R\) is the permutation that interchanges the upper angles and the lower angles. \(P_{TR}\) is the composite permutation of \(P_T\) and \(P_R\). \(P_{TR}\) may also be interpreted as the rotation of a chromosome image \(180°\) about an axis perpendicular to its own plane.

Denote these four permutations as the set \(G\),

\[
G = \{P_0, P_T, P_R, P_{TR}\}.
\]

Lemma 5. Given a chromosome image \(A\), \(\mu_s\) is invariant over \(G\).

\[
\mu_s(A) = \mu_s(g(A)), \quad g \in G.
\]

Lemma 6. The set \(G\) over \(A\) forms an Abelian group.

Definition 3. Given two chromosome images \(A\) and \(B\), \(A\) is angularly equal to \(B\) denoted as \(A = B\) if and only if \(A_i = B_i\) for \(1 \leq i \leq 8\).

Definition 4. Given two chromosome images \(A\) and \(B\), \(A\) is angularly equivalent to \(B\) denoted as \(A = g (B)\) for some \(g\) in \(G\).

Example 8. Given chromosome images \(A\) and \(B\) with

\[
A = (180°, 0°, 180°, 0°, 180°, 0°, 180°, 0°)
\]

\[
B = (0°, 180°, 0°, 180°, 0°, 180°, 0°, 180°)
\]

then \(A = B\).

Theorem 2. There are six most unsymmetrical chromosome images, namely

\[
A^1 = (180°, 0°, 180°, 0°, 180°, 0°, 180°, 0°)
\]

\[
A^2 = (180°, 0°, 180°, 0°, 180°, 0°, 180°, 0°)
\]

\[
A^3 = (180°, 0°, 180°, 0°, 180°, 0°, 180°, 0°)
\]

\[
A^4 = (180°, 0°, 180°, 0°, 180°, 0°, 180°, 0°)
\]

\[
A^5 = (180°, 0°, 180°, 0°, 180°, 0°, 180°, 0°)
\]

\[
A^6 = (180°, 0°, 180°, 0°, 180°, 0°, 180°, 0°).
\]

Proof. \(\mu_s(A) = 0\) implies that \(\sum_{i=1}^{4} |A_{2i-1} - A_{2i}| = 720°\). This implies that

\[
|A_{2i-1} - A_{2i}| = 180°, \quad 1 \leq i \leq 4.
\]

Therefore, either

\[
A_{2i-1} = 0° \text{ and } A_{2i} = 180°
\]

or

\[
A_{2i-1} = 180° \text{ and } A_{2i} = 0°, \text{ for } 1 \leq i \leq 4.
\]

Thus, there are 16 most unsymmetrical chromosome images not angularly equal to each other. By using the equivalence relation defined in Definition 4, these 16 most unsymmetrical chromosome images are degenerated into six equivalent classes, and \(A^1, A^2, A^3, A^4, A^5, \text{ and } A^6\) are the representatives of these six equivalent classes.

E.4 Median Images

When classifying chromosome images, different types and sizes are encountered. Three of these different types, median, submedian, and acrocentric will be examined. They all have the same general appearance. The difference occurs in the location of their centromere. Both figures and string representations of median, submedian, and acrocentric chromosome images are given by Lee and Fu. It was proved that the triangle most dissimilar to a right triangle is \((180°, 0°, 0°)\). Therefore, either

\[
A_{2i-1} = 0° \text{ and } A_{2i} = 180°
\]

or

\[
A_{2i-1} = 180° \text{ and } A_{2i} = 0°, \text{ for } 1 \leq i \leq 4.
\]

Thus, there are 16 most unsymmetrical chromosome images not angularly equal to each other. By using the equivalence relation defined in Definition 4, these 16 most unsymmetrical chromosome images are degenerated into six equivalent classes, and \(A^1, A^2, A^3, A^4, A^5, \text{ and } A^6\) are the representatives of these six equivalent classes.
Lemma 7. Given a chromosome image \( A \) with angles \( A_i \) and sides \( a_j \), the following three conditions are equivalent:

1. \( \mu_M (A) = 1. \)
2. \( \mu_M^* (A) = 1. \)
3. Chromosome image \( A \) is a symmetrical chromosome image and \( a_1 = a_2 = a_3 = a_4. \)

Lemma 8. Given a chromosome image \( A \),
\[
\mu_M (A) \leq \mu_M^* (A).
\]

Lemma 9. Given a chromosome image \( A \), let
\[
\mu_{DM} (A) = 1 - \frac{|a_1 - a_4| + |a_2 - a_3|}{a_1 + a_2 + a_3 + a_4 + a_5}.
\]
Then,
\[
0 \leq \mu_M (A) \leq \mu_M^* (A) \leq \mu_D (A) \leq 1
\]
and
\[
0 \leq \mu_M (A) \leq \mu_M^* (A) \leq \mu_{DM} (A) \leq 1.
\]

Lemma 10. Given a chromosome image \( A \), let
\[
L_M = \{0, \mu_M (A), \mu_M^* (A), \mu_D (A), \mu_{DM} (A), 1\};
\]
then the set \( L_M \) with max and min as the two binary operations forms a distributive but not complemented lattice.

E.5 Submedian Images

By using the terminal set \( V_T \), string representation of a submedian chromosome image is as follows:
\[
X_{\text{submedian}} = cbabbdbbbaabbbcbbbabbbdbbbab.
\]

Given a chromosome image \( A \) with angles \( A_i \) and sides \( a_j \), let
\[
a_{SM} = \min \{ |a_1 - 2a_4| + |a_2 - 2a_3|, |2a_1 - a_4| + |2a_3 - a_5| \}
\]
and
\[
a_T = a_1 + a_2 + a_3 + a_4 + a_5.
\]
A quantitative measure of the similarity of this chromosome image \( A \) to submedian chromosome images may be defined as:
\[
\mu_{SM} (A) = \mu_D (A) [1 - a_{SM}/2a_T]
\]
or
\[
\mu_{SM}^* (A) = 1 - \mu_D (A) (a_{SM}/2a_T).
\]

Lemma 11. Given a chromosome image \( A \), the following three conditions are equivalent:

1. \( \mu_{SM} (A) = 1. \)
2. \( \mu_{SM}^* (A) = 1. \)
3. Chromosome image \( A \) is a symmetrical chromosome image and (either \( a_1 = 2a_4, a_2 = a_3 \) or \( a_4 = 2a_1, a_3 = 2a_2 \)).

Lemma 12. Given a chromosome image \( A \),
\[
\mu_{SM} (A) \leq \mu_{SM}^* (A).
\]

Lemma 13. Given a chromosome image \( A \), let
\[
\mu_{DSM} (A) = 1 - a_T/2a_T;
\]
then
\[
0 \leq \mu_{SM} (A) \leq \mu_{SM}^* (A) \leq \mu_D (A) \leq 1
\]
and
\[
0 \leq \mu_{SM} (A) \leq \mu_{SM}^* (A) \leq \mu_{DSM} (A) \leq 1.
\]

Lemma 14. Given a chromosome image \( A \), let
\[
L_{SM} = \{0, \mu_{SM} (A), \mu_{SM}^* (A), \mu_D (A), \mu_{DSM} (A), 1\};
\]
This set \( L_{SM} \) with max and min as the two binary operations forms a distributive but not complemented lattice.
Depending on its prospective application, the constant 2 used in defining \( a_{SM} \) may be changed to other constants.

E.6 Acrocentric Images

Given the terminal set \( V_T \), string representation of an acrocentric chromosome image is as follows:
\[
X_{\text{acrocentric}} = cadbbbbbaabbbcbbbbabbbabbbba.
\]

Given a chromosome image \( A \) with angles \( A_i \) and sides \( a_j \), let
\[
a_{AC} = \min \{ |a_1 - 4a_4| + |a_2 - 4a_3|, |4a_1 - a_4| + |4a_2 - a_3| \}
\]
then a quantitative measure of the similarity of this chromosome image \( A \) to acrocentric chromosome images may be defined as
\[
\mu_{AC} (A) = \mu_D (A) \cdot \left[1 - a_{AC}/4a_T\right]
\]
or
\[
\mu_{AC}^* (A) = 1 - \mu_D (A) \cdot (a_{AC}/4a_T).
\]

Lemma 15. Given a chromosome image \( A \), the following three conditions are equivalent:
1. \( \mu_{AC}(A) = 1 \).
2. \( \mu_{\tilde{AC}}(A) = 1 \).
3. Chromosome image \( A \) is a symmetrical chromosome image and (either \( a_1 = 4a_4, a_2 = 4a_3 \) or \( a_4 = 4a_1, a_3 = 4a_2 \)).

**Lemma 16.** Given a chromosome image \( A \),

\[
\mu_{AC}(A) \leq \mu_{\tilde{AC}}(A).
\]

**Lemma 17.** Given a chromosome image \( A \), let

\[
\mu_{DAC}(A) = 1 - \frac{a_{AC}}{4a_{T}}.
\]

Then

\[
0 \leq \mu_{AC}(A) \leq \mu_{\tilde{AC}}(A) \leq \mu_{DAC}(A) \leq 1
\]

and

\[
0 \leq \mu_{AC}(A) \leq \mu_{\tilde{AC}}(A) \leq \mu_{DAC}(A) \leq 1.
\]

**Lemma 18.** Given a chromosome image \( A \), let

\[
L_{AC} = \{0, \mu_{AC}(A), \mu_{\tilde{AC}}(A), \mu_{DAC}(A), 1\};
\]

then the set \( L_{AC} \) with max and min as the two binary operations forms a distributive but not complemented lattice.

Depending on its prospective application, the constant 4 used in defining \( a_{AC} \) may be changed to other constants.

**E.7 An Algorithm**

Algorithms for classifying a triangle as an “approximate isosceles triangle,” “approximate equilateral triangle,” “approximate right triangle,” “approximate isosceles right triangle,” or “ordinary triangle,” and algorithms used to classify a quadrangle as “approximate square,” “approximate rectangle,” “approximate rhombus,” “approximate parallelogram,” “approximate trapezoid,” or “ordinary quadrangle” have been presented elsewhere. This section describes an algorithm for classifying a chromosome image into one of the following three classes: “approximate median,” “approximate submedian,” or “approximate acrocentric.”

This can be done in the following manner. Compute \( \mu_M, \mu_{SM}, \) and \( \mu_{AC} \), and set a threshold \( \delta \), where \( \delta \) is a parameter and \( 0 \leq \delta < 1 \). If we compare the max \( \mu_M, \mu_{SM}, \mu_{AC} \) with \( \delta \), there are two possibilities:

1. If \( \max \{ \mu_M, \mu_{SM}, \mu_{AC} \} > \delta \), then there are two possibilities:
   a. If the maximum is unique, then we choose the class corresponding to the maximum value and classify the image accordingly.
   b. If the maximum is not unique, then we define a priority among \( \mu_M, \mu_{SM}, \) and \( \mu_{AC} \) and classify the chromosome image accordingly.
2. Otherwise, the chromosome image is rejected as not belonging to any of the classes.

Depending on its prospective application, \( \mu_M, \mu_{SM}, \) and \( \mu_{AC} \) may be replaced by \( \mu_{M'}, \mu_{S'M'}, \mu_{AC} \), respectively.

**E.8 Expansion or Contraction Constant \( \sigma \)**

The expansion or contraction of a chromosome image will not affect the values of the 8 angles. Let \( \sigma \) be an expansion or contraction constant, with \( \sigma > 0 \). By applying an expansion or contraction constant \( \sigma \) to a chromosome image \( A \) with angles \( A_i \) and sides \( a_j \), we obtain a new chromosome image denoted as \( \sigma A \), with angles \( A_i \) and sides \( \sigma a_j \).

**Lemma 19.** \( \mu_a \) is invariant with respect to an expansion or contraction constant \( \sigma \):

\[
\mu_a(\sigma A) = \mu_a(A).
\]

**Definition 5.** A quantitative measure of a chromosome image is a shape-oriented measure if and only if this measure is invariant with respect to an expansion or contraction constant \( \sigma \).

**Example 9.** \( \mu_a \) is a shape-oriented measure.

**Lemma 20.** \( \mu_M, \mu_{M'}, \mu_{SM}, \mu_{SM'}, \mu_{DSM}, \mu_{DSM'}, \mu_{AC}, \mu_{AC}' \) and \( \mu_{DAC} \) are shape-oriented measures, and are independent of the size of chromosome images.

**Lemma 21.** The classification algorithm described in Section E.7 is independent of the size of chromosome images, and is a shape-oriented classification algorithm.

**E.9 Conclusion of Section E**

A shape-oriented classification algorithm has been described. As we have stated elsewhere, there are three advantages of shape-oriented similarity measures. These are:

1. Two chromosome images may have the same shape but differ in area and dimensions and still be similar.
2. Shape-oriented similarity measures can be normalized between 0 and 1.
3. Shape-oriented similarity measures are invariant with respect to rotation, translation, or expansion or contraction in size.

As Rock demonstrated, the perception of form embodies the automatic assignment of a top, a bottom, and sides. Thus, orientation plays an important role in unassisted chromosome classification. Owing to the invariance of angle and length measurements with respect to orientation, shape-oriented chromosome classification would not be confused by the orientation of the chromosome. In this sense, shape-oriented chromosome classification is better than unassisted chromosome classification. The results obtained in this section may have useful applications in the storage, retrieval, and classification of chromosome images and geometric figures. In addition, the results may be applied to other areas, for instance,
to the work done by Pavlidis" in the area of approximating an arbitrary shape by polygons, the work done by Dacey in the development of a two-dimensional language that produces line pictures of polygons, or the work presented by Harmon on face recognition. The results may also be of use in pattern recognition, information retrieval and artificial intelligence.

We conclude this section with another proposition. (As shown in Figure 3(d), \((A_1, A_2)\) and \((A_3, A_4)\) are exterior biangles; \((A_5, A_6)\) and \((A_7, A_8)\) are interior biangles.)

**Proposition 5.** For shape-oriented storage of chromosome images, if we logically order all the chromosome images individually and independently according to the angular sums of exterior biangles and interior biangles, then we can reduce retrieval time for answering queries such as "retrieve the chromosomes that are very very similar to a given chromosome \(A\)."

**F. DECISION SUPPORT SYSTEMS INVOLVING LEUKOCYTE IMAGES**

Let the universe of discourse \(U_L = \{L\}\) be the universe of leukocyte images. Let the set of terms \(T_L\) be leukocytes with circular nuclear shape, leukocytes with elongated nuclear shape, leukocytes with spiculated nuclear shape, leukocytes with indented nuclear shape, leukocytes with slightly indented nuclear shape, and leukocytes with deeply indented nuclear shape.

The peripheral blood leukocytes have been classified into eight categories. The categories are small lymphocytes, medium lymphocytes, large lymphocytes, band neutrophils, segmented neutrophils, eosinophils, basophils, and monocytes. Pictures of typical peripheral blood leukocytes are given by Bacus and Gose. The features they used were nuclear size, nuclear shape, nuclear cytoplasmic texture, cytoplasm color, and cytoplasm colored texture.

As Bacus and Gose stated, it is particularly important not only to classify the five major cell types, but also to determine "intraspecies differences, between younger and older cells in some classes." For example, intraclass percentage shifts relate to the production rates and maturation of new cells, and thus to the physiological response to stress. In this section, our attention will be focused on the study of the shape properties of leukocytes in order to determine intraclass differences.

Young has stated that the nuclear shapes of lymphocytes, monocytes, eosinophils and basophils are round, indented, segmented, two-lobed and elongated, respectively. Four shape features (circular, spiculated, oblong and irregular) were listed by Bacus and Aggarwal as typical features for computerized microscopic-image analysis; however, they did not show how to measure these four shape features. The nuclear-shape measure used by Bacus and Gose was \((\text{perimeter})^2/\text{area}\). Let \(f\) denote this measure. Then,

\[ f = (\text{perimeter})^2/\text{area}. \]

Bacus and Gose give a way to obtain perimeter and area measure. Out of all possible nuclear shapes, circular shapes minimize \(f\); the minimum of \(f\) is equal to \(4\pi\).

**F.1 Equal-Perimeter Circular-Shape Measure**

The range of \(f\) is from \(4\pi\) to infinite. In order to normalize this measure between 0 and 1, a normalization constant \(p_1\) is multiplied times the reciprocal of \(f\):

\[ \mu_1 = p_1/f \]

where \(p_1\) is a normalization constant, and is set to \(4\pi\). \(\mu_1\) may be viewed as a quantitative measure of the meaning of an "approximately circular shape" or the meaning of an "approximately round shape."

**F.2 Equal-Area Circular-Shape Measure**

Let \(P\) and \(A\) denote, respectively, the perimeter and the area of a nucleus. Let \(P'\) denote the perimeter of a circle with area \(A\). Let \(A'\) denote the area of a circle with perimeter \(P\). Depending on its prospective application, the meaning of an "approximately circular shape" may also be expressed as:

\[ \mu_2 = p_2P'/P, \]

where \(p_2\) is a normalization constant, and is set to 1 in order to normalize the value of \(\mu_2\) between 0 and 1. \(\mu_2\) may be called an equal-perimeter circular-shape measure and \(\mu_2\) an equal-area circular-shape measure. The relationship between \(\mu_1\) and \(\mu_2\) is summarized in the following three theorems.

**Theorem 3.** For all possible nuclear shapes, the equal-perimeter shape measure \(\mu_1\) is always less than or equal to the equal-area circular-shape measure \(\mu_2\).

**Theorem 4.** For all possible nuclear shapes, the equal-perimeter circular-shape measure \(\mu_1\) is always equal to the square of the equal-area circular-shape measure \(\mu_2\).

**Theorem 5.** For any two nuclear shapes \(A\) and \(B\), \(\mu_1(A)\) is greater than or equal to \(\mu_1(B)\) if and only if \(\mu_2(A)\) is greater than or equal to \(\mu_2(B)\).

The proofs of theorems can be obtained from the author. Theorem 5 shows the order-preserving property of \(\mu_1\) and \(\mu_2\) with respect to all possible nuclear shapes.

**F.3 Elongated-Shape Measures**

The nuclear shape of the basophil cell type is elongated. A quantitative measure of the elongated visual concept of a nucleus may be defined as follows:

1. Determine its area and center of mass. A way to determine its area and center of mass is given by Bacus and Gose.
2. Find the best-fit rectangle with the same area and with its center coinciding with the center of mass of the nucleus.

3. Let $a$ and $b$ denote the sides of the best-fit rectangle. Let $\mu_e'$ denote a quantitative measure of the elongated visual concept of the nucleus. Then

$$\mu_e' = \frac{\max(a, b)}{a + b}.$$

4. If the best-fit rectangle is not unique, the $\mu_e$ is defined to be the maximum of $\mu_e'$:

$$\mu_e = \max(\mu_e').$$

F.4 Spiculed-Shape Measures

A quantitative measure of the spiculed visual concept of a nucleus may be expressed as

$$\mu_{sp} = \mu_{e}.$$

Depending on its prospective application, the exponent used in defining $\mu_{sp}$ may take on different values.

F.5 Indented-Shape Measures

The nuclear shape of the monocyte cell type is indented. A quantitative measure of the visual concept "indented," as used of a nucleus, may be defined as follows:

1. As shown in Figure 4, determine points $A$ and $B$ such that the indented nucleus is symmetric with respect to $AB$. If the symmetric axis $AB$ does not exist, determine the axis $AB$ that will minimize the symmetrical difference.

2. Determine points $C$ and $D$ such that the tangents at $C$ and $D$ are perpendicular to the tangent at $A$. If $C$ is not unique, then determine the middle point as $C$.

3. Determine points $E$ and $F$ such that $AE$ is equal to $EC$, and $AF$ is equal to $FD$ along the perimeter.

4. Determine $\theta_1$, which is the angle formed by the tangents at points $E$ and $F$. $\theta_1$ may be called the exterior angle of an indented nucleus.

5. Determine points $G$ and $H$ such that the tangents at $G$ and $H$ are parallel to the tangent at $A$. If $G$ is not unique, then determine the middle point as $G$.

6. Determine points $I$ and $J$ such that $BI$ is equal to $IG$, and $BJ$ is equal to $IH$ along the perimeter.

7. Determine $\theta_2$, which is the angle formed by the tangents at points $I$ and $J$. $\theta_2$ may be called the interior angle of an indented nucleus.

8. Let $\mu_i$ denote a quantitative measure of the indented visual concept of the nucleus. Then,

$$\mu_i = 1 - \rho_i \min(\theta_1, \theta_2)$$

or

$$\mu_i' = 1 - \rho_i' \max(\theta_1, \theta_2)$$

or

$$\mu_i'' = 1 - \rho_i'' \theta_1 \theta_2,$$

where $\rho_i$, $\rho_i'$, and $\rho_i''$ are normalization constants, set to $1/180^\circ$, $1/180^\circ$ and $1/(180^\circ)^2$, respectively, in order to normalize the values of $\mu_i$, $\mu_i'$, and $\mu_i''$ between 0 and 1.

Example 10

1. For Figure 5(a), $\theta_1 = \theta_2 = 0^\circ$. Thus,

$$\mu_i = 1.$$

2. For Figure 5(b), $\theta_1 = 25^\circ$ and $\theta_2 = 10^\circ$. Thus,

$$\mu_i = 0.86.$$

3. For Figure 5(c), $\theta_1 = 84^\circ$ and $\theta_2 = 70^\circ$. Thus,

$$\mu_i = 0.53.$$

4. For Figure 5(d), $\theta_1 = 180^\circ$ and $\theta_2 = 180^\circ$. Thus,

$$\mu_i = 0.$$

5. For Figure 5(e), $\theta_1 = 90^\circ$ and $\theta_2 = 180^\circ$. Thus,

$$\mu_i = 0.$$

It is of interest to note that $\mu_i$ is monotone decreasing with respect to the exterior angle $\theta_1$ and the interior angle $\theta_2$.

F.6 Slightly Indented Shape Measures

A quantitative measure of the slightly indented visual concept of a nucleus may be expressed as

$$\mu_{si} = \mu_{i}^{1/2}.$$

Example 11. (1) For Figure 5(b), $\mu_{si} = 0.93$. (2) For Figure 5(c), $\mu_{si} = 0.73$. 
Figure 5—The grade membership of indented nuclei

Depending on its prospective application, the exponent used in defining \( \mu_{si} \) may take on different values.

### F.7 Deeply Indented Shape Measures

A quantitative measure of the visual concept “deeply indented,” for a nucleus, may be expressed as

\[
\mu_{di} = \mu_1^2
\]

**Example 12.** (1) For Figure 5(b), \( \mu_{di} = 0.74 \). (2) For Figure 5(c), \( \mu_{di} = 0.28 \).

Depending on its prospective application, the exponent used in defining \( \mu_{di} \) may take on different values.

**Proposition 6.** For shape-oriented storage of leukocytes, it is suggested that leukocytes be logically ordered individually and independently according to the magnitudes of the grade of membership of approximately circular (round) nuclei, elongated nuclei, indented nuclei, slightly indented nuclei, and deeply indented nuclei, so that, for answering queries, the amount of leukocyte data that must be searched can be reduced and the response time improved.

### G. CONCLUSIONS

The foregoing analysis has shown that the concepts of fuzzy languages and pictorial databases can be applied to decision support system design methodologies. Decision support systems involving geometric figures, chromosome images, or leukocyte images are presented as illustrative examples.

In similarity retrieval from a pictorial database, very often it is desired to find pictures (or feature vectors, histograms, etc.) that are most similar to or most dissimilar to a test picture (or feature vector). Algorithms for finding most similar images may be found elsewhere. The unique pair of the most dissimilar chromosome images is shown in Figure 6.

Earlier, a review and evaluation of the state of the art of similarity retrieval and updating techniques was presented; various types of similarity measures for extracting picture primitives were described. Using similarity measures, one can not only store similar pictures logically or physically close to each other to improve retrieval efficiency, one can also use such similarity measures to answer fuzzy queries involving nonexact retrieval conditions.

The applications of fuzzy languages and pictorial databases to decision support system design methodologies offer what appears to be a fertile field for further study. The underlying ideas are interesting and easy to apply practically. The results have useful applications in decision support systems, pattern recognition, pictorial information systems and artificial intelligence.

Figure 6—The unique pair of the most dissimilar chromosome images

### REFERENCES
