A New Pattern Representation Scheme Using Data Compression

Presented by

Chen-hsiu Huang

August 17, 2002

Media Data Processing

- How to deal with tremendous variety of media:
  - Caterization resolves the latent global information structure contained in a set of unknown data
  - Recognition provides the means of correctly identifying unknown data

Media Analysis Scheme

- We believe that a new media analysis scheme should alleviate following:
  1. Generality, i.e., applicability to media data of any type
  2. Facility for both categorization and recognition
  3. The ability to cope with indefinitely varying (difficult to represent by a set of finite well-defined models) media data
  4. Easily implementable and low processing cost.
- We will examine a number of candidate general schemes later

VQ: Vector Quantization

- VQ is applicable to a wide range media data and is implemented very easily.
- Categorization and recognition are performed by partitioning a given feature space into several classes and assigning an unknown vector to an appropriate class
- Due to the lack of general mapping schemes, VQ has been limited to rather low-level analysis tasks for which intrinsic feature vectors are available

Frequency Domain and NN: Neural Networks

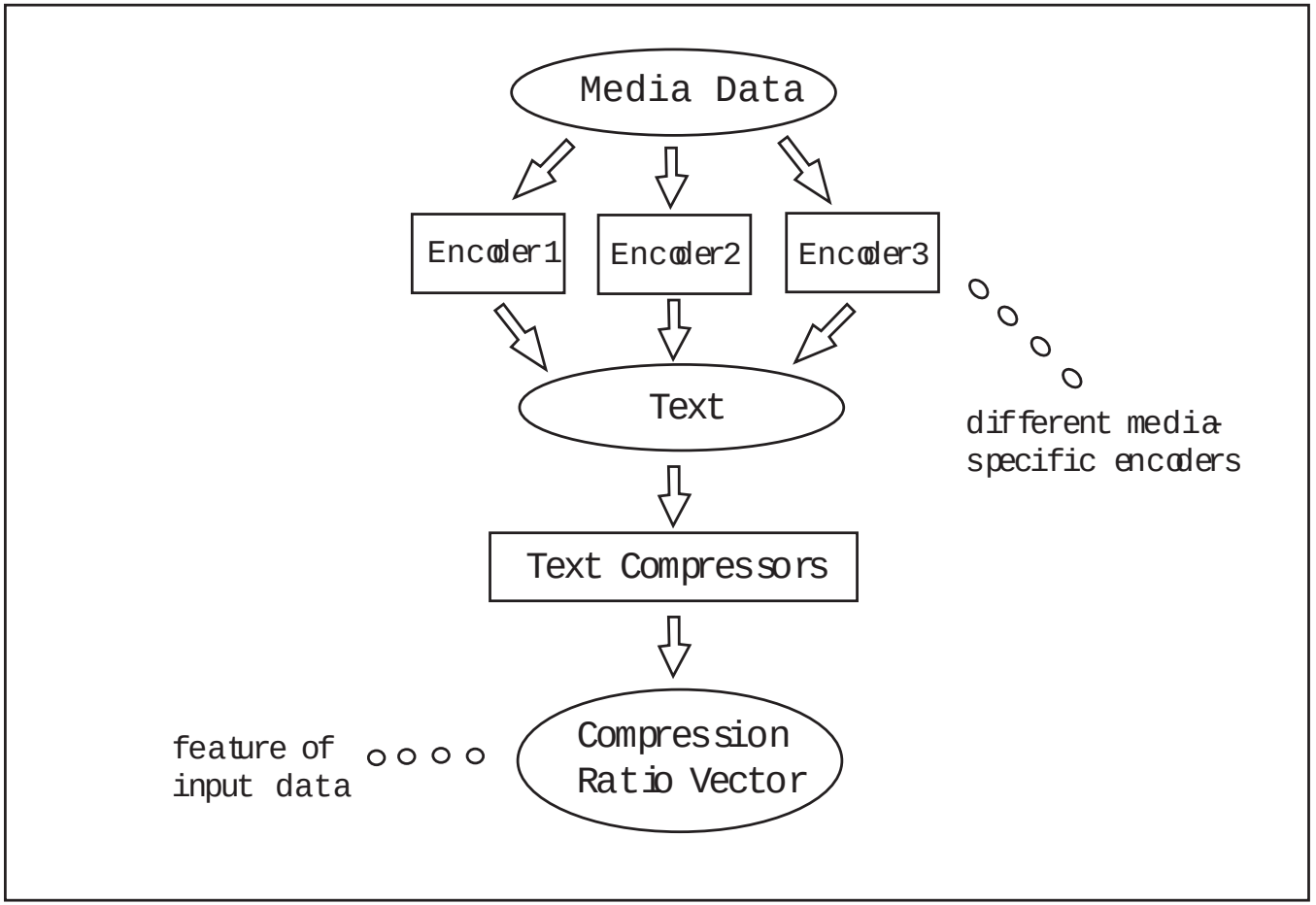
- Frequency domain methods using features such as Fourier-, DCT-, or wavelet-coefficients have wide applicability but require real-valued inputs. This requirement restricts the applicability of such methods to symbol data such as text
- NN and related algorithms provide a completely different general scheme. Such systems also cope well with indefinitely varying sources, yet are only applicable to recognition and require much training even for small tasks with a corresponding high processing cost

The PRDC System

- In PRDC, input data is converted into text and compressed using a set of encoding dictionaries; it generates a compression ratio vector (CV) as a feature of the original input
- The CV is then used as a feature vector in traditional VQ. Although some of the original information is lost in the text generation, we can still exploit the attractive properties of VQ by delimiting its scope

- The realization of PRDC depends on the ability to convert various media data into text and construct a CV feature sapce
- Methods for evaluating the complexity or randomness of finite sequences have been studied extensively providing the foundation for a number of data compression techniques
- However, the use of a CV as a general pattern feature, the core of PRDC, is a new concept

PRDC Diagram



Featuring a Text by CV

- Let $A = \{a_i | 0 \leq i \leq n-1\}$ be an alphabet composed of n characters. A text t is a finite sequence over A. Let $l(t)$ be its length
- For example, if $A = \{a, b\}$, then $t_1 = aaaaaa, t_2 = aabaab, t_3 = ababab, t_4 = abbabb$ and $t_5 = bbbbbb$ are texts, and $l(t_i) = 6$ for all i
- Concatenation of two texts, u and v, is denoted uv. Note that $l(uv) = l(u) + l(v)$

- Call substring of t as *words*, and define a dictionary $d_{m,t}$ as a set of words in t with $l(t) \leq m$. For example,
  $d_{3,t_2} = \{a, b, aa, ab, ba, aab, aba, baa\}$ and
  $d_{3,t_4} = \{a, b, ab, ba, bb, abb, bab, bba\}$
- A parsing of u by $d_{m,t}$, which is denoted by $p(u, d_{m,t})$ is a successive partitioning of u into words of $d_{m,t}$
- If $d_{m,t}$ contains a significant amount of information about u, the parsed word count $l(p(u, d_{m,t}))$ is small. Note that $p(u, d_{m,t})$ is not unique

- In order to ensure the uniqueness of parsing, we introduce the concept of greedy parsing $gp(u, d_{m,t})$
- This is defined as the recursive parsing of a text u by taking the longest prefix $lpf(u, d_{m,t})$ of u in $d_{m,t}$ followed by the greedy parsing of the remaining part $rest(u)$, as defined by the following function (ϕ denotes null text)

$$gp(u, d_{m,t}) = \begin{cases} \phi & \text{if } u = \phi \\ lpf(u, d_{m,t}).gp(rest(u), d_{m,t}) & \text{otherwise} \end{cases}$$

- For example, $gp(t_1, d_{3,t_4}) = a.a.a.a.a.a$, $l(gp(t_1, d_{3,t_4})) = 6$ and $gp(t_3, d_{3,t_4}) = ab.ab.ab$, $l(gp(t_3, d_{3,t_4})) = 3$. The uniqueness of $gp(u, d_{m,t})$ is proven in Theorem 1.
- Now, we can define the compression ratio $\rho(u, d_{m,t})$ of u by $d_{m,t}$

$$\rho(u, d_{m,t}) = \frac{l(gp(u, d_{m,t}))}{l(u)}$$

- Using the above example, we get $\rho(t_1, d_{3,t_4}) = (6/6) = 1.0$ and $\rho(t_3, d_{3,t_4}) = (3/6) = 0.5$
- $0.5 < 1.0$ means t_3 resembles t_4 rather t_1 resembles t_4

- In order to enhance the featuring power, let us use a tuple of dictionaries $D_{m,t} = (d_{m,t_1}, d_{m,t_2}, ..., d_{m,t_n})$ construct from a text set $T = \{t_1, t_2, ...t_n\}$. Then we can define an n-dimensional CV of u.

$$\vec{\rho}(u, D_{m,T}) = (\rho(u, d_{m,t_1}), ..., \rho(u, d_{m,t_n}))$$

- If we choose $T = \{t_2, t_4\}$ and $m = 3$, we obtain the CVs shown in Table 1.
- The distance between these vectors represent similarities between the original texts.

CVs for Example Texts

- Table 1.

Text\Function	$gp(t_i, d_{3,t_2})$	$gp(t_i, d_{3,t_4})$	$l(gp(t_i, d_{3,t_2}))$	$l(gp(t_i, d_{3,t_4}))$	$\vec{p}(u, D_{3,\{t_2,t_4\}})$
$t_1 = aaaaaa$	$aa.aa.aa$	$a.a.a.a.a.a$	3	6	$(0.5, 1.0)$
$t_2 = aabaab$	$aab.aab$	$a.ab.a.ab$	2	4	$(0.33, 0.67)$
$t_3 = ababab$	$aba.ba.b$	$ab.ab.ab$	3	3	$(0.5, 0.5)$
$t_4 = abbabb$	$ab.ba.b.b$	$abb.abb$	4	2	$(0.67, 0.33)$
$t_5 = bbbbbb$	$b.b.b.b.b.b$	$bb.bb.bb$	6	3	$(1.0, 0.5)$

Mathematical Discussion of CVs

- For CVs to be valid feature vectors of texts, the following minimum requirements must be met:
  1. The CV of any text t must be able to be determined uniquely
  2. Similmarities between texts should be adequately reflected in the CVs

- For (1), we show in Theorem 1 that the use of greedy parsing allows us to map a text to a unique CV in a multidimensional unit cube spanned by $D_{m,T}$
- As for (2), we point out in Theorem 2 that the mapping from a text to its CV may be degenerative, i.e., different texts can be mapped to an identical CV
- However, in Theorem 3, we show that we can remedy this situation by extending $D_{m,T}$. We show in Theorems 4 and 5 that similar texts are mapped to similar CVs.

Mathematical Proofs

Theorem 1. *For any text u and dictionary tuple*

$$D_{m,T} = (d_{m,t_1}, d_{m,t_2}, ..., d_{m,t_n}),$$

the compression ratio vector $\vec{\rho}(u, D_{m,T})$ is determined uniquely. Moreover, $\vec{\rho}(u, D_{m,T}) \in [0,1]^{|T|}$, where $|T|$ is the cardinality of T and $[0,1]^{|T|}$ is a $|T|$-dimensional unit cube

Proof. The uniqueness of $\rho(u, d_{m,t_k}) = (1/l(u))l(gp(u, d_{m,t_k}))$ follows from the uniqueness of $l(u)$ and $l(gp(u, d_{m,t_k}))$. As the former is obvious, we show the latter by showing its minimality. Suppose contrarily that some parsing $p(u, d_{m,t_k}) < l(gp(u, d_{m,t_k}))$, then for some i, the ith word w_{p_i} of $p(u, d_{m,t_k})$ should leave the ith word w_{gp_i} of $gp(u, d_{m,t_k})$ behind, giving $w_{p_i} = u' w_{gp_i} v' \in d_{m,t_k}$ for some $v' \neq \phi$ (u' might be ϕ). This means that $w_{gp_i} v' \in d_{m,t_k}$ by the definition of

d_{m,t_k}, contradicting the greediness of w_{gp_i}. Therefore, $l(gp(u, d_{m,t_k}))$ should be minimal. The latter part of the theorem follows from the obvious fact $1 \leq l(gp(u, d_{m,t_k})) \leq l(u)$ and the definition of $\rho(u, d_{m,t_k})$. □

Theorem 2. *Let u and v be texts, then*

$$u = v \Rightarrow \vec{\rho}(u, D_{m,T}) = \vec{\rho}(v, D_{m,T}),$$

but the converse is not always true.

Proof. The first part of the Theorem is obvious. The last part is shown by counter example. Let $A = \{a, b, 0, 1\}$, $u = ababbb$, $v = 010111$, $T = \{ababbb010111, 010111ababbb\}$ and $D_{2,T} = (\{a, b, ab, ba, bb, b0, 0, 1, 01, 10, 11\}, \{0, 1, 01, 10, 11, 1a, a, b, ab, ba, bb\})$

We then have $\vec{\rho}(u, D_{2,T}) = \vec{\rho}(v, D_{m,T}) = (0.5, 0.5)$. But, $u \neq v$. □

$\Rightarrow$ This problem can be patched by adding the responsible texts to the dictionary tuple T.

Theorem 3. *If*

$$(\vec{\rho}(u, D_{m,T}) = \vec{\rho}(v, D_{m,T})) \wedge u \neq v$$

then

$$\exists \hat{m}.[\vec{\rho}(u, D_{\hat{m},T\cup\{u,v\}}) \neq \vec{\rho}(v, D_{\hat{m},T\cup\{u,v\}})]$$

Proof. Assuming contrarily, let us attempt to refute the conclusion, getting

$$\forall \hat{m}.[\vec{\rho}(u, D_{\hat{m},T\cup\{u,v\}}) = \vec{\rho}(v, D_{\hat{m},T\cup\{u,v\}})]$$

Seperating the compression operations of T and $\{u, v\}$, we get

$$\forall \hat{m}.[\vec{\rho}(u, D_{\hat{m},T}) = \vec{\rho}(v, D_{\hat{m},T}) \wedge \vec{\rho}(u, D_{\hat{m},\{u,v\}}) = \vec{\rho}(v, D_{\hat{m},\{u,v\}})]$$

Using the second term, we get

$$\forall \hat{m}.[\rho(u, d_{\hat{m},u}) = \rho(v, d_{\hat{m},u}) \wedge \rho(u, d_{\hat{m},v}) = \rho(v, d_{\hat{m},v})]$$

Without the loss of generality, suppose either $l(u) > l(v)$ or

$l(u) = l(v)$.

In the case of $l(u) > l(v)$, if we choose

$$\hat{m} = \max(l(u), l(v)) = l(u)$$

we get

$$\rho(u, d_{\hat{m},u}) = 1/l(u) < 1/l(v) \leq \rho(v, d_{\hat{,}u})$$

This contradicts the above formula.

In the case of $l(u) = l(v)$, if we choose $\hat{m} = l(u) = l(v)$ and use the above formula, we obtain

$$\rho(u, d_{\hat{m},u}) = 1/l(u) = 1/lv(v) = \rho(v, d_{\hat{m},u})$$

This means v can be parsed by only one word u, i.e., $u = v$. This contradicts the premise $u \neq v$ □

Finally, we show that similar texts are mapped to similar CVs.

We first show in Theorem 4 that the CV of a concatenated text uv can be approximated by a weighted sum of CVs of u and v.

Then, using this result, we show that a minor variant of u is mapped to a minor variant of the CV of u.

Theorem 4. $\forall u \forall v \forall m \in \{m | m < \min(l(u), l(v))\}$.

$$\left[\left| \vec{\rho}_{uv} - (\frac{l(u)}{l(uv)} \vec{\rho}_u + \frac{l(v)}{l(uv)} \vec{\rho}_v) \right| \leq \frac{r(T)}{l(uv)} \right]$$

where $\vec{\rho}_{uv}$ *abbreviates* $\vec{\rho}(uv, D_{m,T})$*, etc., and* $r(T)$ *is the radius of a unit sphere in* $|T|$*-dimensional space such as* $r(T) = \sqrt{|T|}$ *(Euclidian distance) or* $r(T) = |T|$ *(City distance).*

Moreover, if $l(uvw)$ *is large and* $l(v) \ll l(uvw)$*, that is, if* v *is much shorter than* uvw*, then we get*

$$\vec{\rho}(uvw, D_{m,T}) \approx \vec{\rho}(uw, D_{m,T})$$

Proof. First we proof

$$l(gp(uv, dm, t)) \leq l(gp(u, d_{m,t})) + l(gp(v, d_{m,t})) + 1$$

Suppose the contrary, we get

$$\begin{aligned} l(gp(uv, dm, t)) &> l(gp(u, d_{m,t})) + l(gp(v, d_{m,t})) + 1 \\ &> l(gp(u, d_{m,t})) + l(gp(v, d_{m,t})) \end{aligned}$$

This means it's possible to obtain a trivial nongreedy parsing $p(uv, d_{m,t}) = gp(u, d_{m,t}).gp(v, d_{m,t})$ sufficing $l(p(uv, d_{m,t})) < l(gp(uv, d_{m,t}))$. This contradict Theorem 1.

Second, we prove

$$l(gp(u, d_{m,t})) + l(gp(v, d_{m,t})) - 1 \leq l(gp(uv, dm, t))$$

Assuming that $gp(uv, d_{m,t}) = w_{gp_1}.w_{gp_2}...w_{gp_l}$, then there exists a word w_{gp_k} such that $w_{gp_k} = w^-_{gp_k} w^+_{gp_k}$ and $gp(u, d_{m,t}) = w_{gp_1}.w_{gp_2}...w^-_{gp_k}$. We are given a nongreedy parsing $p(v, d_{m,t}) = w^+_{gp_k}.w_{gp_{k+1}}...w_{gp_l}$ with

$$l(p(v, d_{m,t})) \leq l - k + 1 = l(gp(uv, d_{m,t})) - l(gp(u, d_{m,t})) + 1$$

Here, the inequality ($<$) holds when $w^+ = \phi$ and

$$l(p(v, d_{m,t})) = l(gp(uv, d_{m,t})) - l(gp(u, d_{m,t})) = l - k$$

As $l(gp(v, d_{m,t})) \leq l(p(v, d_{m,t}))$ by Theorem 1, we get

$$l(gp(v, d_{m,t})) \leq l(p(v, d_{m,t})) \leq$$
$$l(gp(uv, d_{m,t})) - l(gp(u, d_{m,t})) + 1$$

This implies

$$l(gp(u, d_{m,t})) + l(gp(v, d_{m,t})) - 1 \leq l(gp(uv, dm, t))$$

Dividing these two inequalities by $l(uv)$ and using $\frac{1}{l(uv)} = \frac{l(u)}{l(uv)} \frac{1}{l(u)}$ and $\frac{1}{l(uv)} = \frac{l(v)}{l(uv)} \frac{1}{l(v)}$, we obtain

$$\left| \rho_{u,v} - \left(\frac{l(u)}{l(uv)} \rho_u + \frac{l(v)}{l(uv)} \rho_v \right) \right| \leq \frac{1}{l(uv)}$$

As this holds for each element of $|T|$-dimemsional $\vec{\rho}_{uv}$, $\vec{\rho}_u$, and $\vec{\rho}_v$, we

get required result.

To proof the later part, let $u' = uv$ and use (1) twice. We then get

$$\begin{aligned}\vec{\rho}_{uvw} = \vec{\rho}_{uvw} &\approx \left(\frac{l(u')}{l(u'w)}\vec{\rho}_{u'} + \frac{l(w)}{l(u'w)}\vec{\rho}_w\right) \\ &\approx \left(\frac{l(u')}{l(u'w)}\left[\frac{l(u)}{l(uv)}\vec{\rho}_u + \frac{l(v)}{l(uv)}\vec{\rho}_v\right] + \frac{l(w)}{l(u'w)}\vec{\rho}_w\right) \\ &= \left(\frac{l(u)}{l(uvw)}\vec{\rho}_u + \frac{l(v)}{l(uvw)}\vec{\rho}_v + \frac{l(w)}{l(uvw)}\vec{\rho}_w\right)\end{aligned}$$

Therefore, when $l(v) \ll l(uvw)$, we get

$$\vec{\rho}_{uvw} \approx \frac{l(u)}{l(uvw)}\vec{\rho}_u + 0 + \frac{l(w)}{l(uvw)}\vec{\rho}_w \approx \vec{\rho}_{uw}$$

□

Encoding Media Data into Text

Sequential Pattern. Given a nontext sequence $s = s_1 s_2 ... s_n$, first segment s to obtain

$$SEG(s) = v_1 v_2 ... v_l$$

Replace each segment by as letter to give a text

$$t = VQ(SEG(s)) = VQ(v_1)VQ(v_2)...VQ(v_l)$$

where $VQ(x)$ is a function (vector quantizer) mapping x onto some applied when each s_i is a vector.

Spatial Pattern. Let $P = \{\vec{\rho}_{i,j} | (i,j) \in I_r \times I_c\}$ be a color image composed of pixels of I_r rows and I_c columns, where $\vec{p}_{i,j}$ denotes the RGB-vector of a pixel (i,j). First, compile P into a nondirected weighted graph $G(P)$ composed of nodes $n_{i,j}$, edges $e_{i,j,k,l}$, and edge weights $w_{i,j,k,l}$.

Here, we define the edge weight as the color difference between two terminal ndoes using an appropriate distance function $d(x,y)$, i.e., $w_{i,j,k,l} = d(\vec{p}_{i,j}, \vec{p}_{k,l})$.

Then, transform $G(P)$ into $MST(G(P))$, the minimum spanning tree of $G(P)$. Starting from a node, at the north-west corner, for example, traverse $MST(G(P))$ in a light-weight-edge-first manner outputing a sequence

$$TRAV(MST(G(P))) = (\vec{p}_{i_1,j_1}, dir_{1,0})(\vec{p}_{i_2,j_2}, dir_{2,1})...(\vec{p}_{i_l,j_l}, dir_{l,l-1})$$

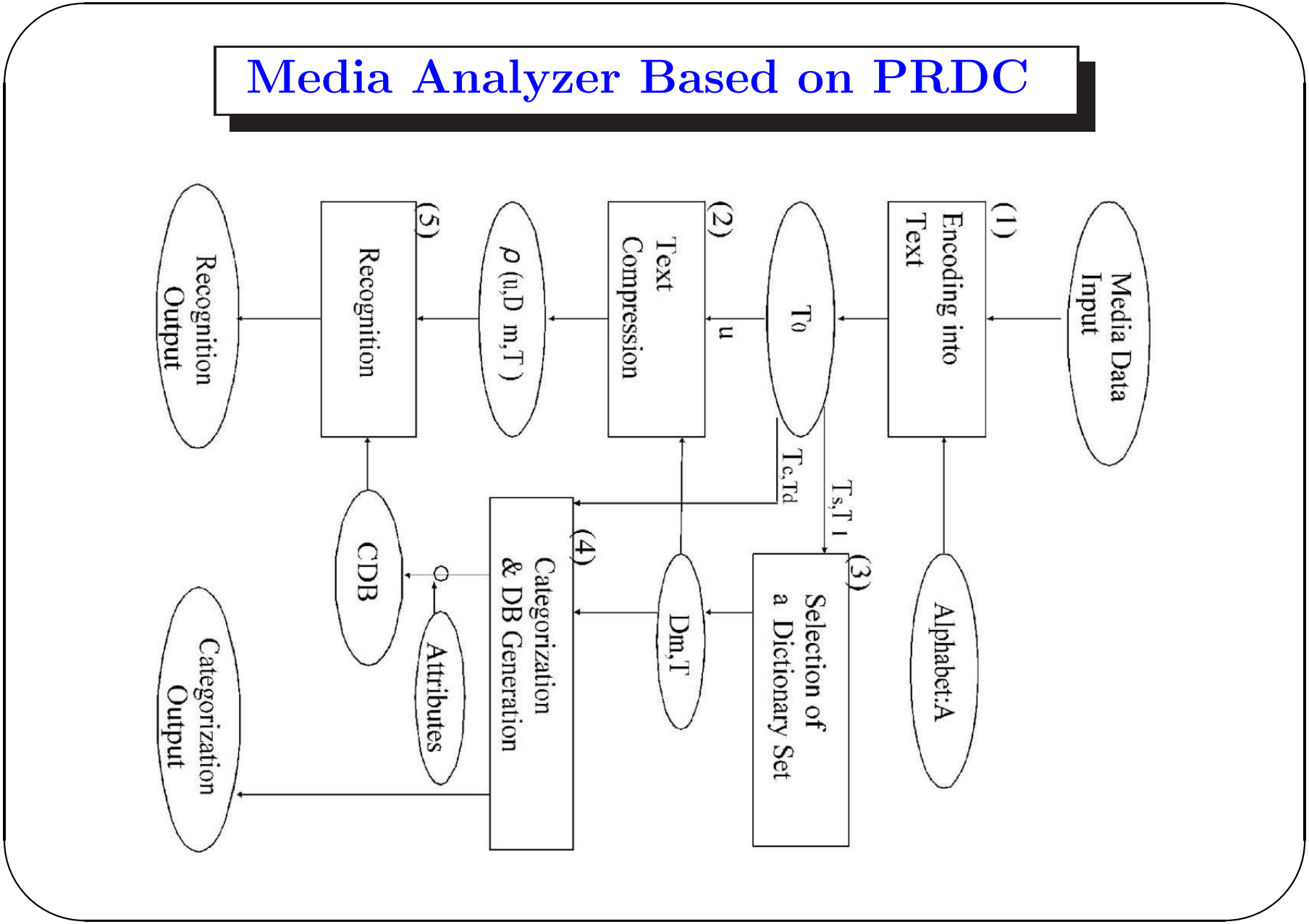
Here, $\vec{p}_{i_k,j_k}$ and $dir_{k,k-1}$ denotes the color of the current node and the traverse direction from the previous node $n_{i_{k-1},j_{k-1}}$ to the current node n_{i_k,j_k}. For example

$$dir_{k,k-1} \in \{horizontal, slant1, vertical, slant2\}$$

Finally, we encode $TRAV(MST(G(P)))$ into a text

$$\begin{aligned} t &= VQ(TRAV(MST(G(P)))) \\ &= VQ((\vec{p}_{i_1,j_1}, dir_{1,0}))...VQ((\vec{p}_{i_l,j_l}, dir_{l,l-1})) \end{aligned}$$

Note that this scheme is an extension of Freeman's chain code [28] in that both color contour shape (part of spatial) information and color (spectral) information on P are encoded simultaneously.



Media Analyzer Based on PRDC

Realizing a Media Analyzer

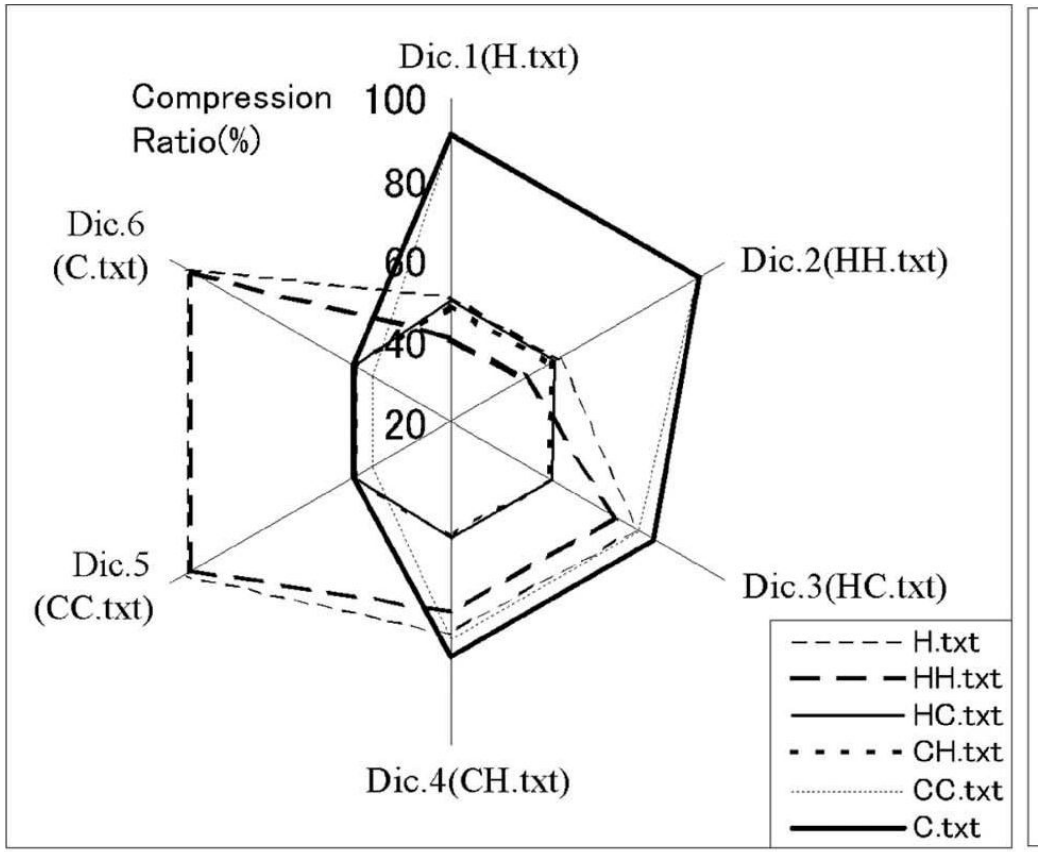
- **Encoding into Text**. Input media data is encoded into text using the method described above.
- **Text Compression**. The buffer-type dictionary approach is adopted in some of the LZ-type text compression algorithms [26], [27].
- **Selection of a Dictionary Set**. First, prepare a small text set T_s to get D_{m,T_s}. Perform cluster analysis on the output vector $\{\vec{\rho}(t, D_{m,T_s}) \in T_l\}$. Set the value $|T|$ equal to the number of clusters found. Choose $|T|$ vectors, one from each cluster. Pick up $|T|$ texts corresponding to there $|T|$ vectors to define $D_{m,T}$.

- **Catergorization and DB Generation**. After determining $D_{m,T}$, we choose a set of training texts T_c and prepare a set of teaching data $\{(v, atr(v))|v \in T_c\}$, where $atr(v)$ is the manually prepared attribute of v. We use $D_{m,T}$ to compress texts $v \in T_c$ to obtain a case database $CDB = \{(\vec{\rho}(v, D_{m,T}), atr(v)|v\ inT_c)\}$. In categorizing a set of texts T_d, we calculate the respective $CVs = \{\vec{\rho}(v, D_{m,T})|v \in T_d\}$, on which we perform a cluster analysis [10].
- **Recognition**. In Recognition tasks, the nearest element $(\vec{\rho}(v^*, D_{m,T}), atr(v^*)) \in CDB$ of the incoming $\vec{\rho}(u, D_{m,T})$ is selected and the corresponding $atr(v^*)$ is output as the recognition result (e.g., the name) for u

Feature Representability of CV

Computer Programs. The header and source parts of C programs are gathered into two files, H.txt and C.txt, both of which contain approximately 2,300 characters. These files are concatenated to form HH.txt, HC.txt, CH.txt, and CC.txt

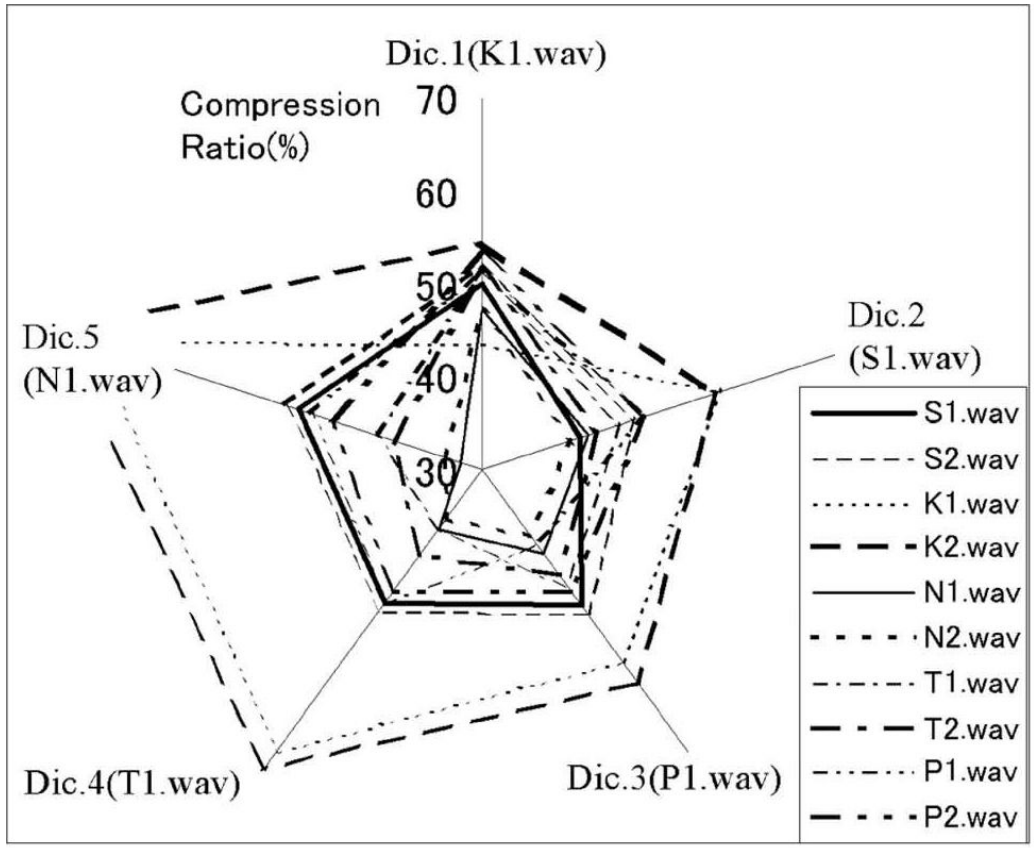
The resolution is quite sharp and six CVs can be clearly categorized into three groups: {H.txt, HH.txt}, {HC.txt, CH.txt}, and {C.txt, CC.txt}. The dictionaries {H.txt, HH.txt} compress {H.txt, HH.txt} and {HC.txt, CH.txt} well, but, as expected, this is not the case for {C.txt, CC.txt}.


Dic.1(H.txt)
Compression Ratio(%)
100
80
60
40
20
Dic.6 (C.txt)
Dic.2(HH.txt)
Dic.5 (CC.txt)
Dic.3(HC.txt)
Dic.4(CH.txt)
H.txt
HH.txt
HC.txt
CH.txt
CC.txt
C.txt

(a)

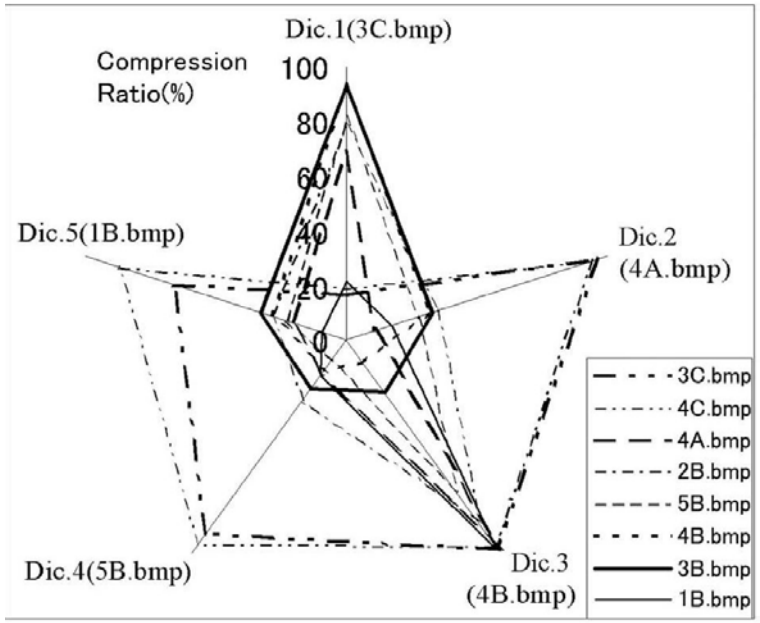
Human Voice. Two 30-second self-introduction speeches are recorded for five students. Each file is divided into frames, and each frame is encoded into one of 26 codes. Several frame lengths were tested and a frame length of 25-ms was found to provide the clearest features

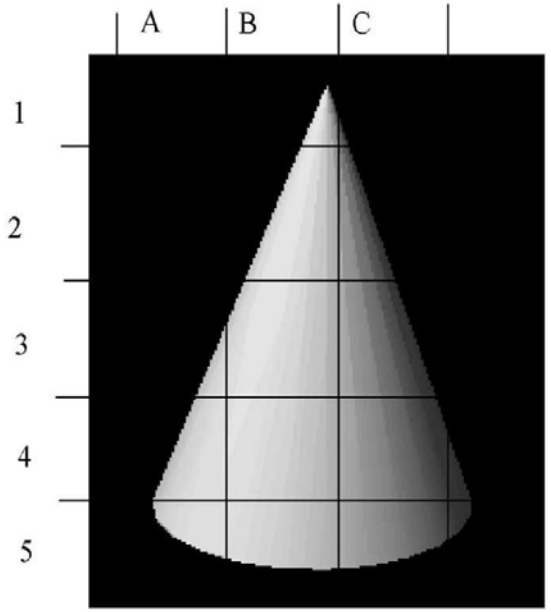
The resolution is not as sharp, except for {K1.wav, K2.wav}. However, it it possible to categorize the data into three groups: {K1.wav, K2.wav}, {S1.wav, S2.wav, P1.wav, P2.wav}, and {T1.wav, T2.wav, N1.wav, N2.wav}.


Dic.1(K1.wav)
Compression Ratio(%)
70
60
50
40
30
Dic.2 (S1.wav)
Dic.3(P1.wav)
Dic.4(T1.wav)
Dic.5 (N1.wav)
S1.wav
S2.wav
K1.wav
K2.wav
N1.wav
N2.wav
T1.wav
T2.wav
P1.wav
P2.wav

Gray-scale Image. Eight areas of 50×50 pixels are selected from Fig. 2d to generate files in bmp format. These subimages are encoded into one of 64 codes, (8 MST directions) $\times$ (8 grayscales), using the proposed spatial pattern coding method.

The resolution is sharp, with four distinct groups: {3C.bmp, 4C.bmp}, {4A.bmp, 2B.bmp, 5B.bmp}, {4B.bmp, 3B.bmp}, and {1B.bmp}. This is in good agreement with our visual impression for the original image. Based onthis categorization, 3B.bmp as an unknown input will be recognizable as being similar to 4B.bmp, in accordance with our intuition.

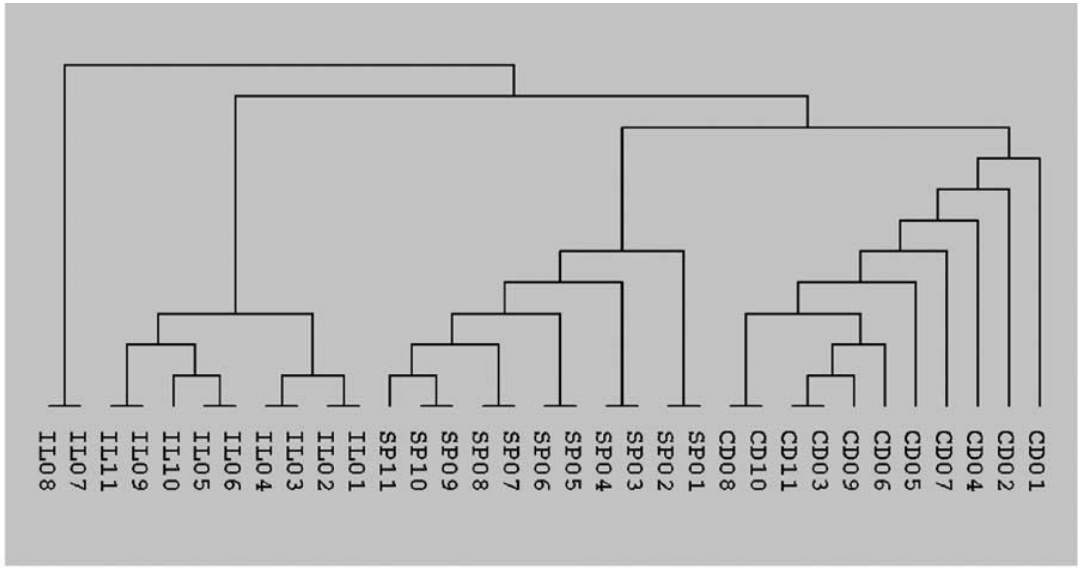

Dic.1(3C.bmp)
Compression Ratio(%)
100
80
60
40
20
0
Dic.5(1B.bmp)
Dic.2 (4A.bmp)
Dic.4(5B.bmp)
Dic.3 (4B.bmp)
3C.bmp
4C.bmp
4A.bmp
2B.bmp
5B.bmp
4B.bmp
3B.bmp
1B.bmp


A
B
C
1
2
3
4
5

(d)

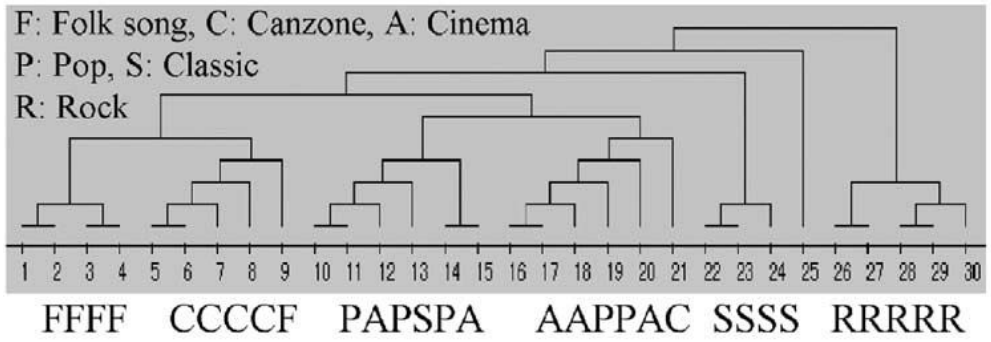
Applications

Genome Categorization (amino acid sequence)

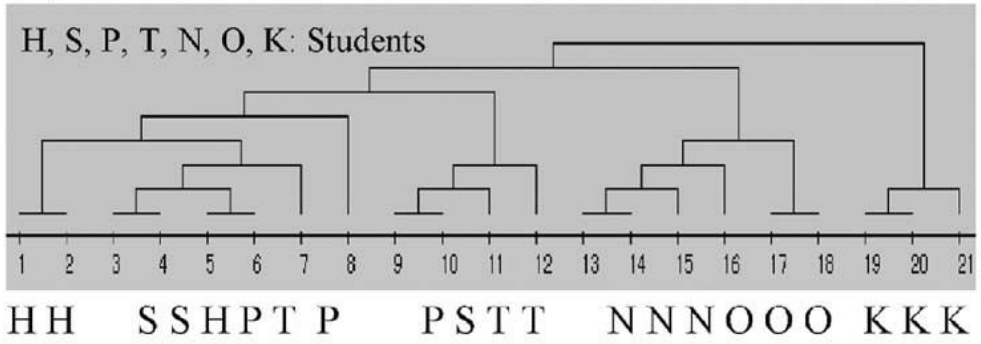


Categorization of Music and Voice

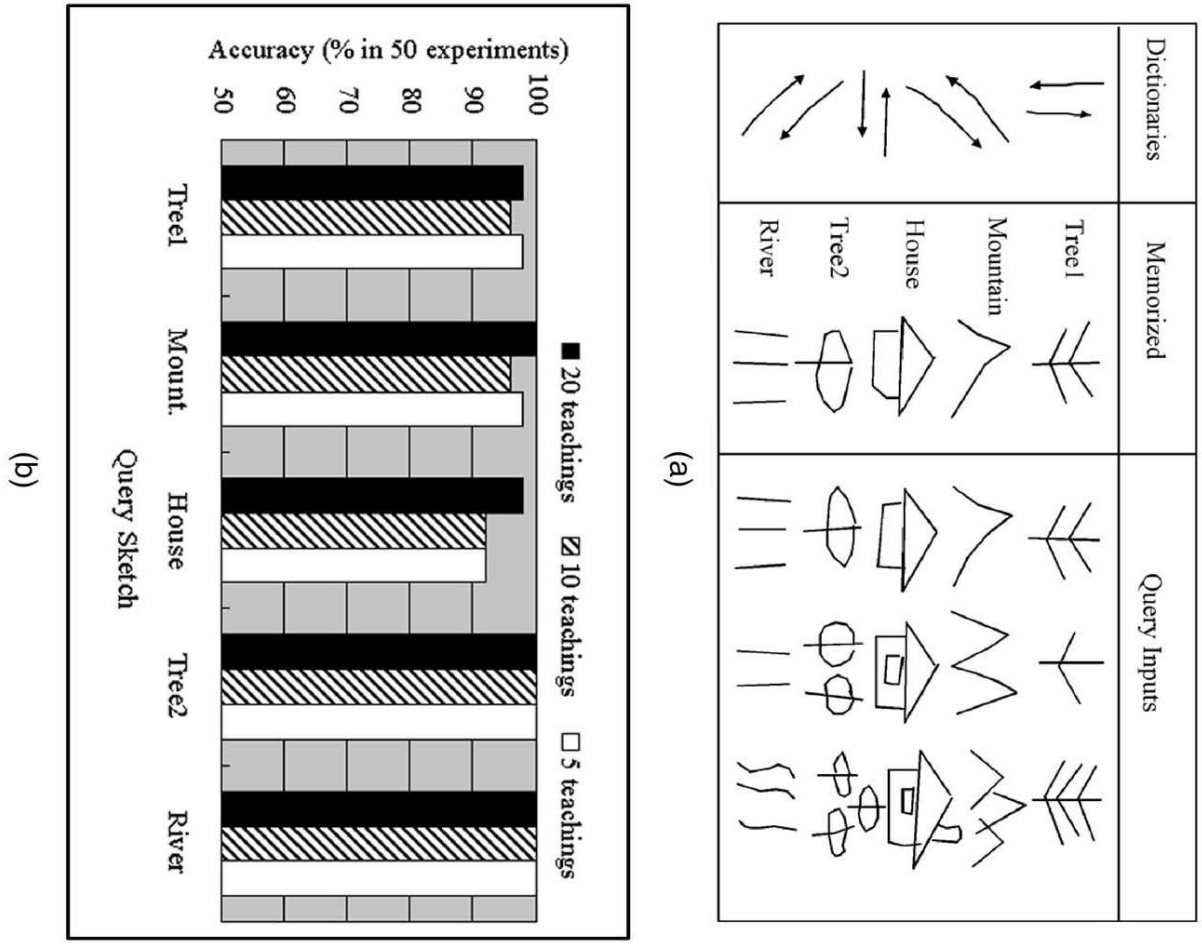
(1) Music: 1 min. 0.5 s/segment



(2) Human Voices: 0.5 min. 0.025 s/segment



Sketch Recognition



Color Image Analysis

Fact\Decision	Sea	Wood	Field	Land	Grass	River	Road	Rail	Building
Sea	100.0								
Wood		83.7	2.0	13.0				1.3	
Field	12.3	3.9	50.8	8.1	24.9				
Land			22.0	31.2			38.5	8.3	
Grass			9.0		55.0			36.0	
River						100.0			
Road				8.3			66.7		25.0
Rail		8.3						91.7	
Building		6.7					30.0		63.3

Confusion Matrix of Land-Cover Recognition. Average accuracy = 74 precent.

Conclusion

- We have proposed a new pattern representation scheme called PRDC, by which input data is converted into a text and then compressed using a set of dictionaries. PRDC can realize attractive properties for media analysis: generality, facility for both categorization (class formation) and recognition (classification), ability to cope with indefinitely varying media data, and easy implementability.
- We have presented a mathematical proof of the realizability of a feature space of CVs and demonstrated the usefulness of PRDC through application to several tasksthat require these properties.

Future Work

- Future investigations include the application of PRDC to more specific and sophisticated media analysis tasks and a performance comparison with other methods.
- We anticipatethat combinations of PRDC with high-level methods will be effective. In addition, we intend to examine a variant of PRDC that uses media-specific compressors rather than universal text compressors. We anticipate that this variant will attain higher analysis power at the expense of generality