

招待講演 2

「画像とメディア統合」

会津大学 コンピュータ理工学部 教授

岡 隆一

<概要>

画像に関わる統合アルゴリズムには、画像内部で機能するものと画像と他のメディアの間で機能するものがある。

前者には、部分と全体の統合の例として、画像全体を局所的な表現に押し込めるモデル(Cellular Feature)を示し、また画像理解のための Pixelと言語の融合の例としては、画像が言語的表現に落ちる(symbol grounding)ためのアルゴリズムを Middle Vision を実現するものとして位置付けたモデルを示した。

後者には、まず、人間と画像を含むメディアの実時間統合を行うもの (strong integration)の例として、人間の思考過程を動画や音声の frame-wise 認識、音声合成、CG 出力などの全てと強く coupling する Human-Computer Interaction のための「実時間完結原理」のアーキテクチャを示した。これによって、Note PC 上でのメディア統合後が容易に実現できるものとなる。次に、画像を含むメディア間での統合には、Bridge Data とよぶものを介する(weak integration)があることを示した。Bridge Data としていくつかの具体的な例を示し、その扱いに適したアルゴリズムの例も示した。

上記の統合アルゴリズムについては、OCR (文字読取装置) やビデオや音楽の検索ソフトとして、一部商用化されたものもあり、そのデモも示した。

結論として、実世界の生データのへ有効性をもつアルゴリズムには、固有のアーキテクチャを有する必要があることを強調した。

画像とメディア統合

岡 隆一

会津大学
コンピュータ理工学部

2003.3.28/R.Oka

1

Outline

1. Integration algorithms for image processing
 - 1.1 Global feature extraction
 - 1.2 Image understanding
2. Integration between image and other media
 - 2.1 Strong integration
 - 2.2 Weak integration

2003.3.28/R.Oka

2

Research activities in the world:

"Integration of Natural Language and Vision Processing (Volume I)
- Computational Model and Systems",
Edited by Paul McKevitt, Kluwer Academic Publishers (1995).
ibid. Volume II — Intelligent Multimedia
ibid. Volume III — Theory and Grounding Representation
ibid. Volume IV — Recent Advances.

"Language, Vision and Music",
Edited by Paul McKevitt, Sean O Nuallain and Conn Mulvihill,
John Benjamins Publishing Company (2002).

2003.3.28/R.Oka

3

1. Integration algorithms for image processing

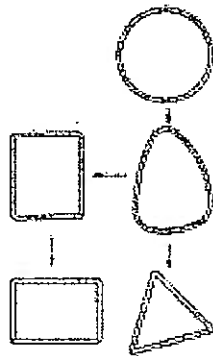
- 1.1 Global feature extraction
- 1.2 Image understanding

2003.3.28/R.Oka

4

1.1 Global feature extraction

A single global feature of image = A whole image

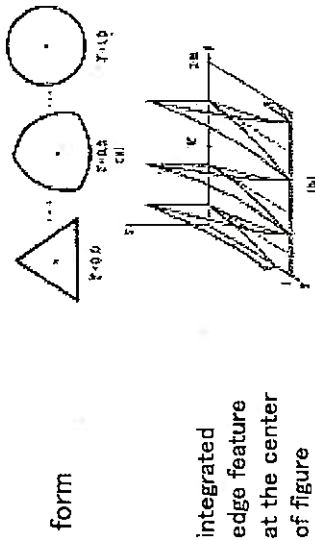


The global feature should be robust for geometrical variations

2003.3.28/R.Oka

5

Form and edge feature



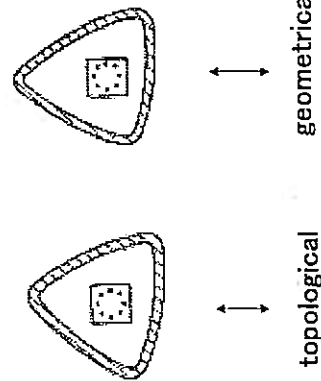
form

integrated edge feature at the center of figure

2003.3.28/R.Oka

6

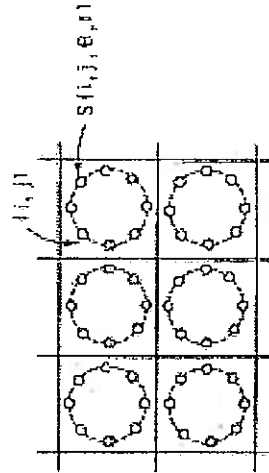
Global feature representation



2003.3.28/R.Oka

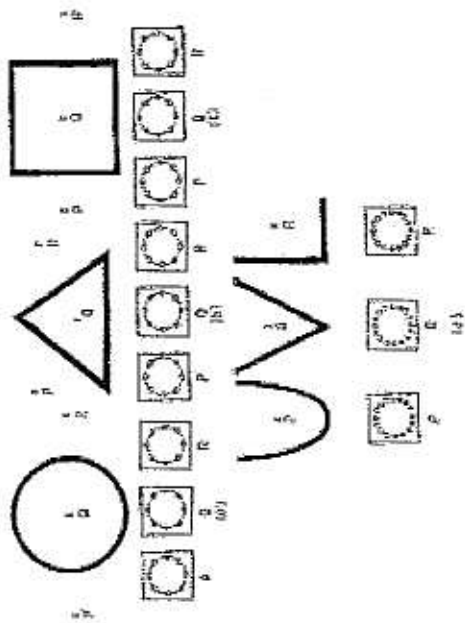
7

Frame work of Cellular Feature (Oka, 1977)



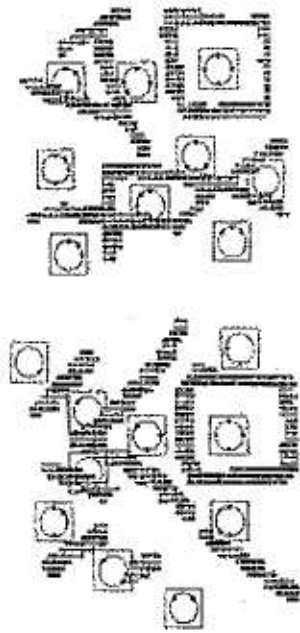
2003.3.28/R.Oka

8



2003.3.28/R_Oka

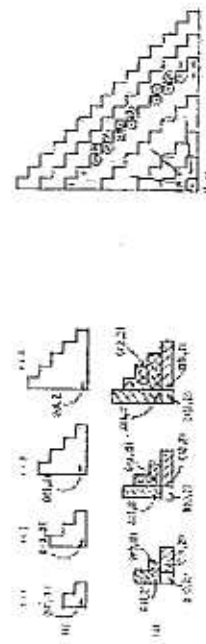
9



Location and its Cellular Feature

2003.3.28/R_Oka

10

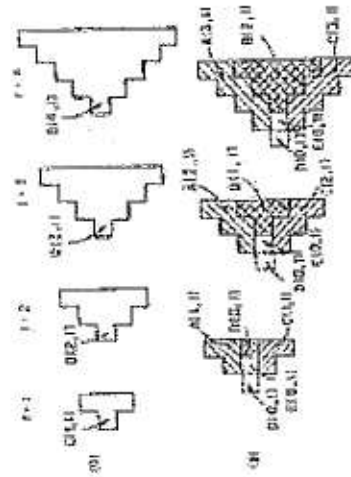


Integration algorithm:

Parallel, local and iterative computation
for integrating edge features on digital plane

2003.3.28/R_Oka

11



2003.3.28/R_Oka

12

1.2 Image understanding



Input image



Output text

A monkey looks at a box

2003.3.28/R.Oka

13

Image Understanding

Early vision (feature extraction etc)

D.Marr etc

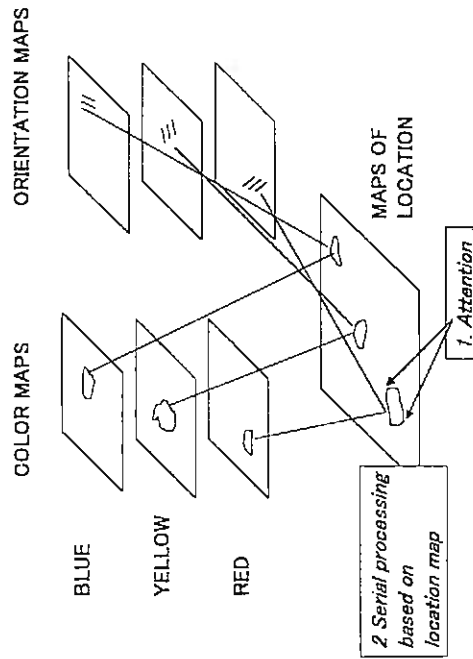
Middle vision — *Cellular Frame Model*
(Oka, 1987)

High level vision (use of knowledge)
Frame Model (M. Minsky etc)

Cognitive Science (Treisman etc)

2003.3.28/R.Oka

14



Treisman Model for feature integration

2003.3.28/R.Oka

15

Multi-layer
space

array

Understanding
Configuration



Each label has a two-dimensional
array. Each location has a value
indicating the existence likelihood
of the label.

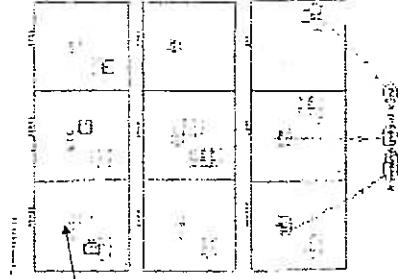


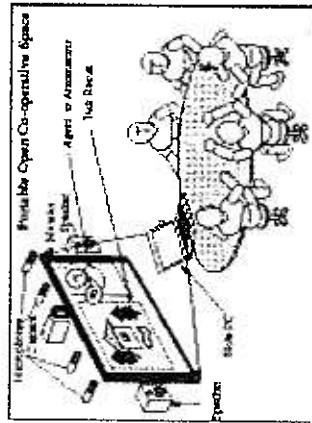
Image understanding based on Cellular Frame Model

2003.3.28/R.Oka

16

2.1 Strong integration

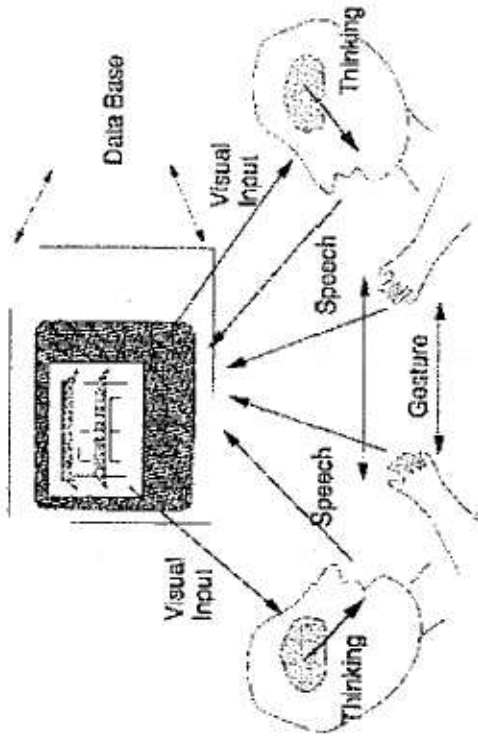
Human Computer Interface



Integration of multimedia and human response

2003.3.28/R.Oka

25



2003.3.28/R.Oka

26

Gesture recognition from a time-varying image

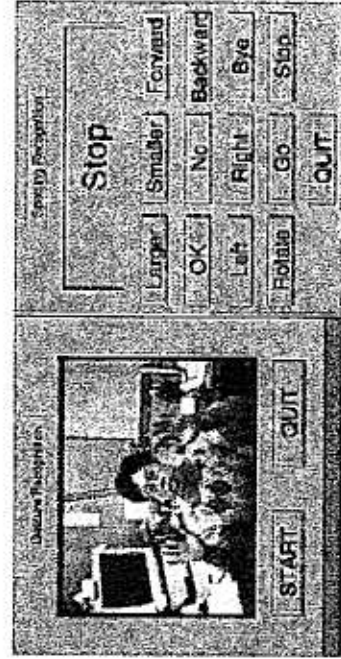
Why gesture ?

1. Typical non-verbal expression
2. New input media

2003.3.28/R.Oka

27

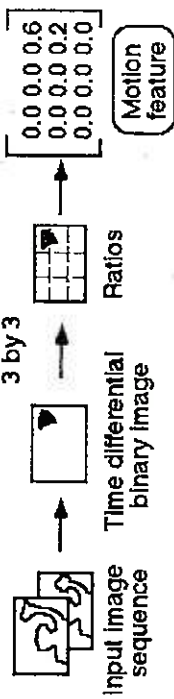
On-line recognition



2003.3.28/R.Oka

28

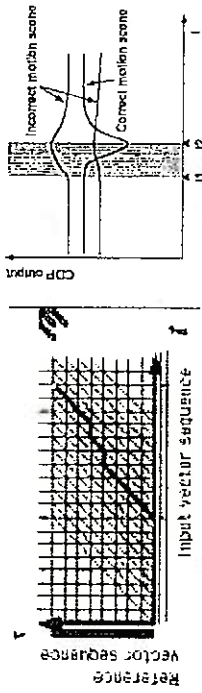
Low-resolution Features



2003.3.28/R.O.km

29

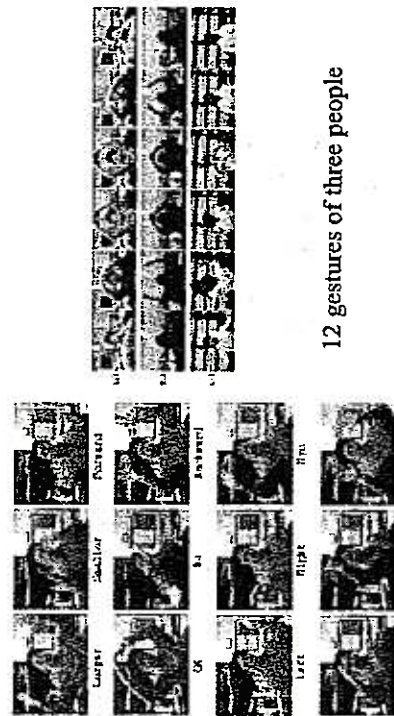
Spotting Recognition with Continuous DP



Realtime and frame-wise recognition without any segmentation

2003.3.28/R.O.km

30



2003.3.28/R.O.km

31

Human side

Query : Speech, Text, Still Image, Humming, Gesture

Evaluation and Thinking

Text, CG Display and Speech Synthesis

Output media

Feature Extraction

Retrieval of Multimedia Data

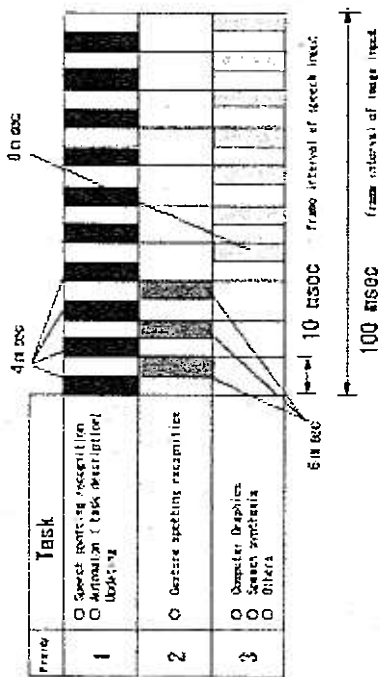
Database side

One loop processing by one frame

2003.3.28/R.O.km Frame-wise Completion Principle

32

Scheduling of computation



2003.3.28/R.Oda

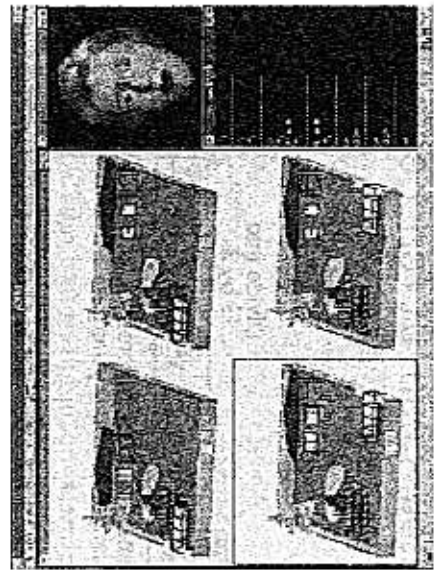
33



Realization of strong integration of multimedia on a notebook PC

2003.3.28/R.Oda

34



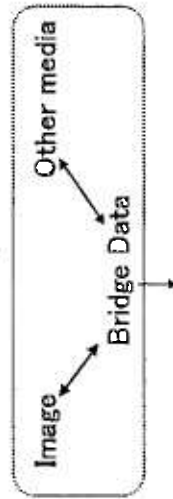
2003.3.28/R.Oda

35

2.2 Weak integration

“Bridge Data” between image and other media

weak integration

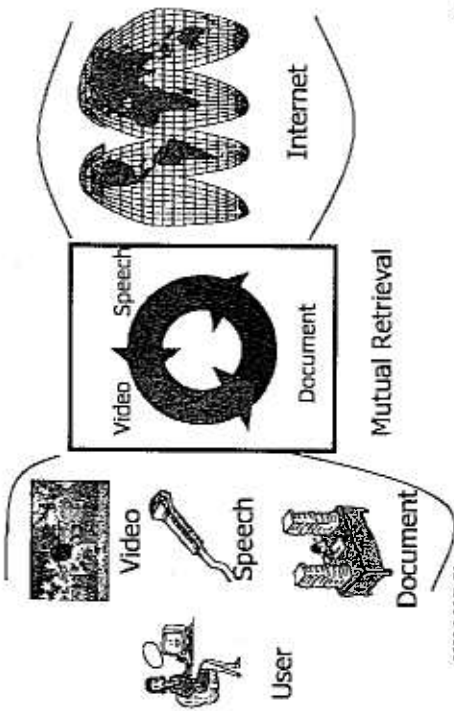


(With/without annotation, MPEG-7, etc.)

2003.3.28/R.Oda

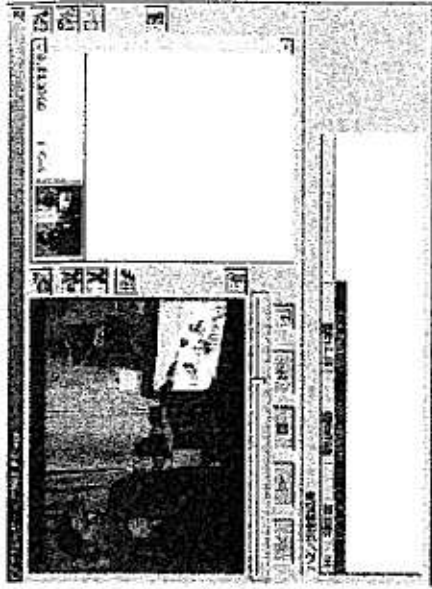
36

CrossMediator (RWC2002)



2003.3.28/R_Oka

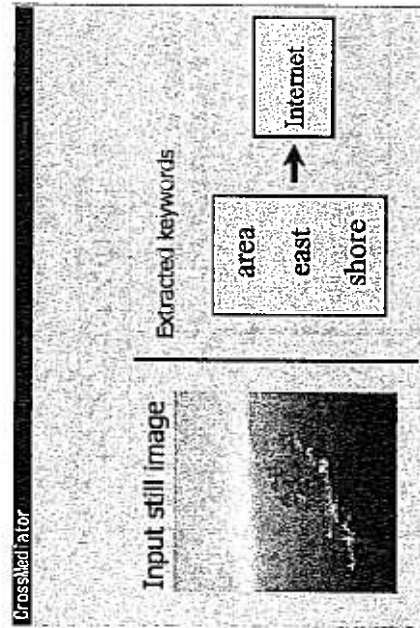
37



Query: humming Output: video scene

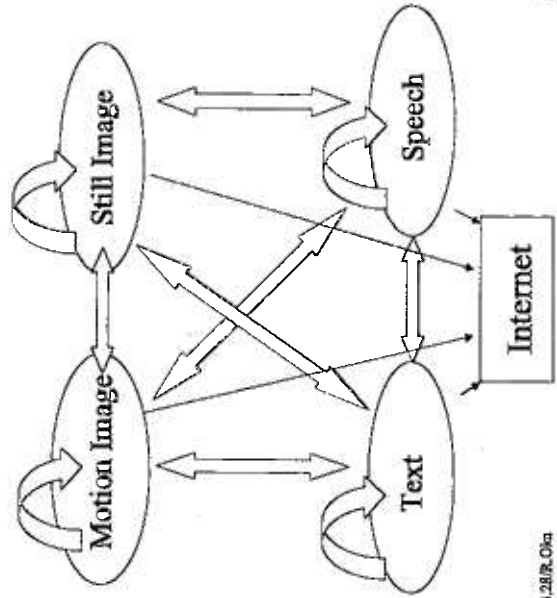
2003.3.28/R_Oka

38



2003.3.28/R_Oka

39



2003.3.28/R_Oka

40

ここでの立場とは、

音声、静止画、動画の信号データの認識
(言語記号化、ただしOCRは除く)を極力避
ける。

なぜか？

実世界で通常にであう信号データを今の
技術では十分には認識できないため。

2003.3.28/R.O.km

45

さらに、

アノテーション作業(MPEG7的方法も含む)も避ける。

なぜか？

大量データについては作業量が膨大となるため。

2003.3.28/R.O.km

46

では、どうする？

メディアを媒介する表現(Bridge Data)と

検索アルゴリズム

をつくる。

2003.3.28/R.O.km

47

- ： 言語による音声波形へのアクセス
- ： 音声による言語データへのアクセス

汎用性、任意の言語への適用

語彙に依存しない

解決法

音声認識を経ない方法
(Bridge Data、Algorithm は?)

2003.3.28/R.O.km

48



4. 多言語対応の必要性
 4.1. 多言語対応の必要性
 4.2. 多言語対応の必要性

採用は日本政府が負担
 チャーシュー機使用も検討



Optical Character Reader

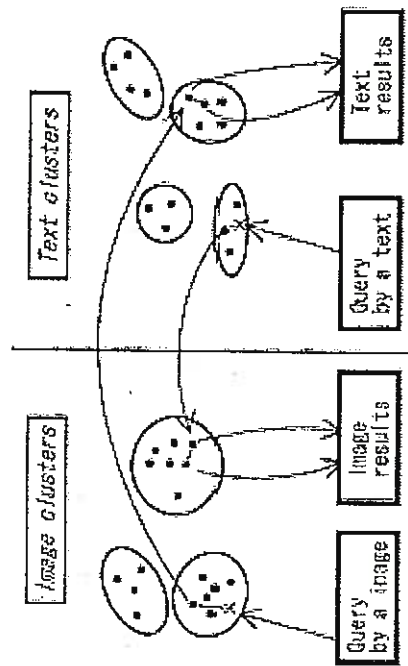
2003.3.28/R.Oka

53

画像とテキストの対データによる 相関データ(Bridge Data)の学習 と検索Algorithm

2003.3.28/R.Oka

54



2003.3.28/R.Oka

55

Pattern raw data processing for retrieval

Soft Indexing by microscopic unit:

- Feature Vector
- VQ code
- Phoneme label
-
- (Non-categorical unit)

Undefined type of database

Simpler and possible to realize automatic indexing by available methods

2003.3.28/R.Oka

Hard Indexing by macroscopic unit:

- Word
- Sentence
- Html description
- Relational Database description
-
- (Categorical unit)

Segmentation finished

Need to solve hard problems so that it becomes difficult to realize automatic indexing

56

A sample speech of topic

• 話題 2.5

小次郎は毎朝パパが店じ家 (みせじじんち) (q25_1)まで送って行く。そして帰りはわたしが迎えに行く。まだ一ヶ月しか経ってないのに不安になる。「こんなん6年間も続ける事が出来るのだからか?」と思う。舎人小 (とねりしょう) (q25_2)にはみどり幼稚園のお友達がたくさん通っている。だから小次郎の為に思っているこの小学校に決めたのだ。親の努力あるのみ!

2003.3.28/R.Okin

57

- ① 自転車で総ス球
- ② 帰りにイトウヨウカドウ
- ③ 遅お向かいにしてしまった
- ④ 幼稚園バスのお迎え
- ⑤ 旅行、久しぶりだけど
- ⑥ 何だか雨っぽい
- ⑦ お産見舞いに行くとぞお
- ⑧ 10時5分竹の塚伊藤邸に着く
- ⑨ 壺穴のジョン
- ⑩ カラオケボックスの前で

データベース中の任意の区間を検索

単語や句や文の単位にこだわらなくてよい

2003.3.28/R.Okin

59

【話題 1】

樹、7時45分学校へ行く。その1時間後、新良がバスに乗って幼稚園へ行く。洗車機を平しでから、総わり決着自転車で総ス球へ10時〜12時までトランポリン、遊りにイトウヨウカドウ立へ行き多額のおがけを返す。公園でたばこを吸うつもりが成がつかず平がつかない。

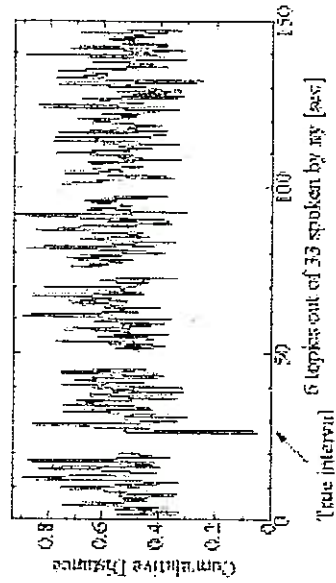
【話題 3】

明日、天気はどうかだろうか。あきって旅行から洗車した1時ほど、互に幼稚園へ行く。車では4・5・6は帰れぬ。遊園、路上で遊びだすと、お兄さんに空うのも一年ぶりがらいかな。土・日仕事だからめつたに急がない。ホテルでソウイーターは休みもらえたんだね。睡しみななな...

2003.3.28/R.Okin

58

A sample of stream of CDP output value



- Query: [マル秘収納テク] including in the first topic.
- Horizontal axis indicates only 6 topics in the 33 topics.

2003.3.28/R.Okin

Word identification from a still image

Internet Retrieval

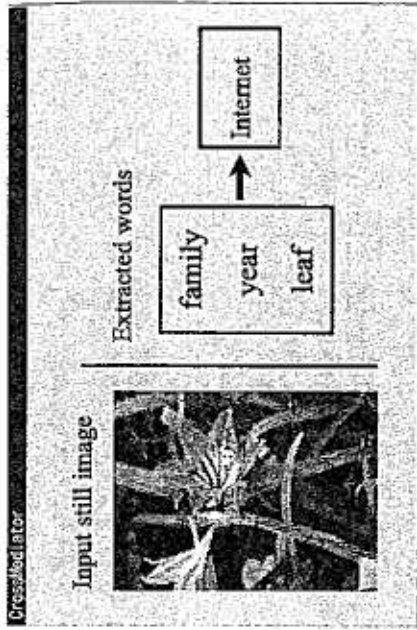
Text query



Image query

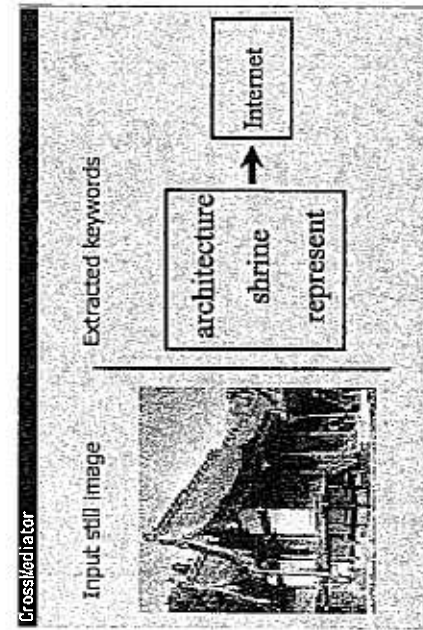
2003.3.28/R.Oka

61



2003.3.28/R.Oka

62



2003.3.28/R.Oka

63

6. Real-time retrieval of video image using a real-time video query

Query: Live TV

Database: Stored Video

2003.3.28/R.Oka

64

Displaying of retrieved data



2003.3.28/R.Oka

70



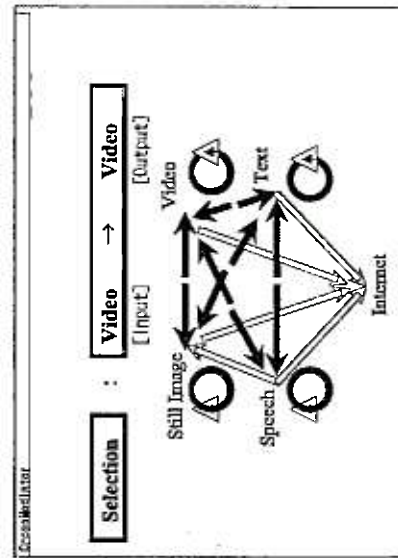
2003.3.28/R.Oka

69



2003.3.28/R.Oka

72

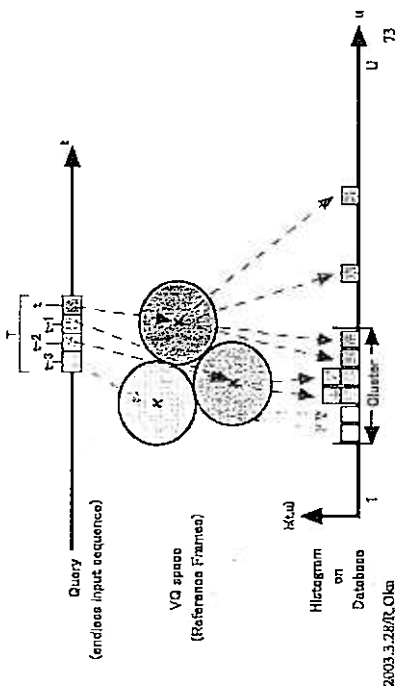


2003.3.28/R.Oka

71

demo

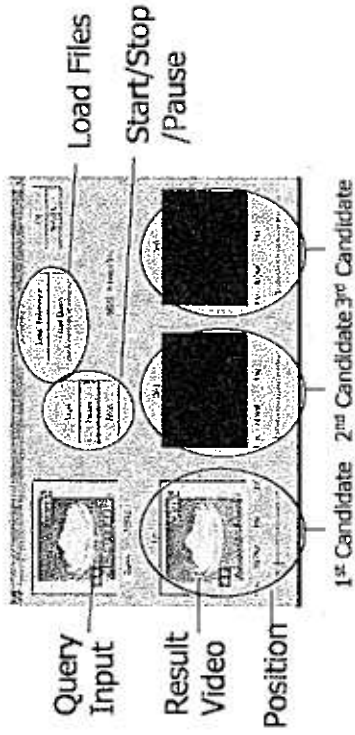
Retrieving Outline(Rutic Method)



2003.3.28/R.O.km

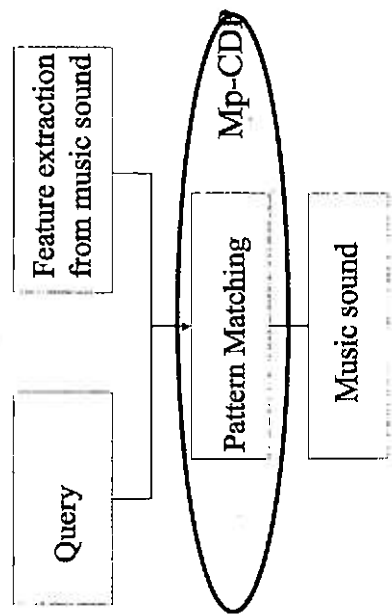
74

Video Chaser Prototype



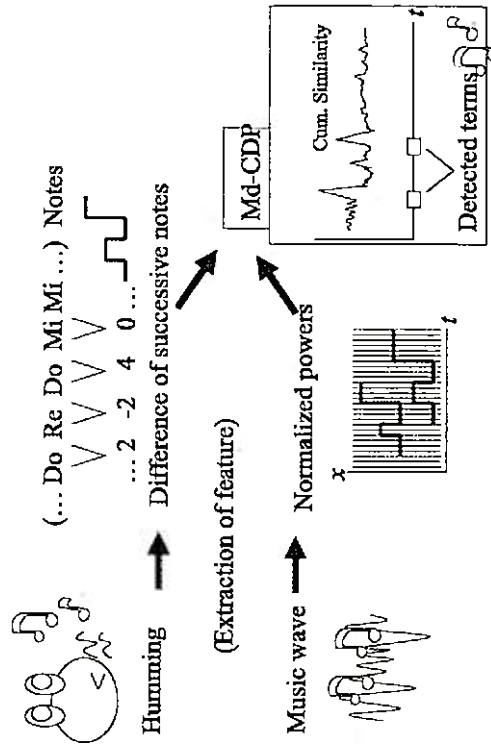
2003.3.28/R.O.km

Outline of music retrieval by humming



2003.3.28/R.O.km

75



2003.3.28/R.O.km

76

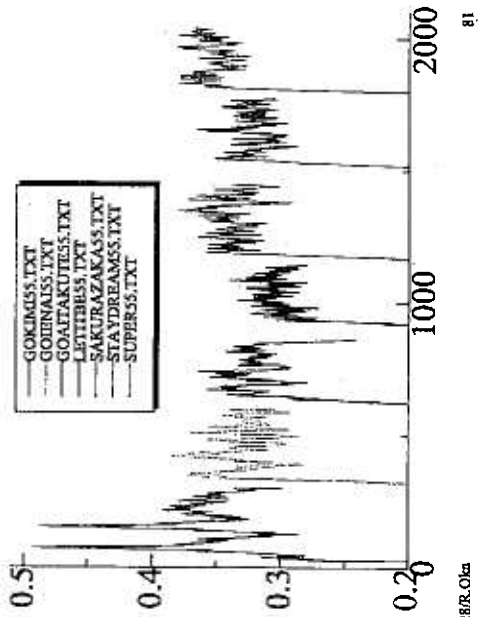
・ 言語による画像(ビデオを含む)データへのアクセス

解決法

1. 言語によってビデオの音声波形へアクセスし、次に時刻を同じくする(Bridge Data)ビデオ画像にアクセスする
2. 画像とテキストの対データによる相関データ(Bridge Data)の言語部分にアクセスし、次に画像データへアクセス

2003.3.28/R.01a

82



2003.3.28/R.01a

81

デモ1 音声によるテキストの検索、さらに音楽も出力

- (1) 歌詞の一部を音声で入力
この一部は歌詞のどんな区間でもOK
- (2) 音声で文字列(歌詞の一部)を検索
- (3) 検索された歌詞の一部を含むフレーズ表示
- (4) (3)の部分の音楽を再生

2003.3.28/R.01a

81

デモ2 音声による音声データの検索

- (1) 過去に録音した音声波形(音声メモ)の任意の部分区間を音声で入力
- (2) 入力音声に類似する音声波形を録音音声から検索
- (3) 類似音声区間を含む音声区間を再生する

2003.3.28/R.01a

84

デモ 3

音声またはテキストによる
ビデオ動画像の検索、

- (1) TV番組を録画する
- (2) 音声またはテキストで
みたいところを入力する
- (3) 録画データの音声部分において
(2)の入力と類似した部分を検索する
- (4) (3)で検索した時刻を含む
ビデオ(動画と音声)区間を再生する

2003.3.28/R.Ola

85

デモ 4

歌声で音楽を検索

- (1) TVの歌や音楽番組、家庭で
歌っているところを録画する
- (2) 再生したい歌や音楽の知って
いる部分を、歌ってまたはハミング
で入力する。入力するものは
歌や音楽の任意の部分でよい。
- (3) 録画データの歌や音楽の部分に
(2)と類似した部分を検索する
- (4) (3)で検索した時刻を含む
ビデオ(動画と歌や音楽)を再生する

2003.3.28/R.Ola

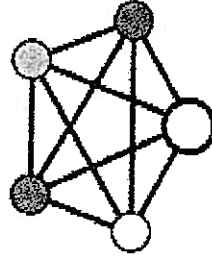
86

Commercial products

1. 1980 A private company developed Optical Character Reader based on my method called "Cellular Feature" and marketed as a character recognition package software. This OCR package software now commands a 60 percent share of the Japanese OCR market.
2. 2000 I proposed a new concept of integrated multi-media retrieval software, incorporated the concept into a specific software called "CrossMediator," and succeeded in marketing the software through a private company. The software was recommended at 2001 Mac World New York as an excellent software and also was demonstrated by Craig (CEO, Intel) in the keynote speech **Oracle OpenWorld 2001 Keynote Craig Barrett, President and CEO, Intel Corporation Moscone Convention Center San Francisco, Calif. December 3, 2001.**
3. 1980-2001 Many private companies including Exing Inc., PATRIS Corp., Hitachi Ltd., Fuji Xerox Co. Ltd., Modivdrive Inc. etc. have developed or are under the development of software based on my works to produce commercial products.

2003.3.28/R.Ola

87



CrossMediator

Real World Computing Partnership
Tsukuba Research Center

2003.3.28/R.Ola

88

Thank you for your attention.

2003.3.28/R.O.ka

89