MVA '92 IAPR Workshop on Machine Vision Applications Dec. 7-9,1992, Tokyo

AN EXPERIMENT OF REALIZATION OF FULLY AUTOMATED KEYWORD EXTRACTION IN IMAGE DATABASE SYSTEMS

Jun Yamane, Masao Sakauchi
Institute of Industrial Science, The University of Tokyo
7-22-1, Roppongi, Minato-ku, Tokyo 106, Japan

ABSTRACT

Recently, a flexible image database retrieval system where image keywords can be captured automatically is strongly required, to manage a practical number of image data successfully. But image recognition/understanding technology level is not generally inefficient enough to achieve this requirement.

In order to overcome this problem, we propose a new image database framework here. In the proposed system, image keywords are extracted in fully-automated fashion by the deviced image recognition system. Image keywords used here is a collection of recognized objects in the image, where recognition levels are allowed to be intermediate or imperfect. We introduce the concept of "recognition thesaurus" for managing these various level keywords successfully. As an embodiment of this concept we implemented an image database with various types of sports scenes. Retrieval evaluations reveals the effectiveness of the proposed method.

1 INTRODUCTION

Recently, a flexible image database retrieval system where image keywords can be captured automatically is strongly required, to manage a practical number of image data successfully. But image recognition/understanding technology level is not generally inefficient enough to achieve this requirement.

In order to overcome this problem, we propose a new image database framework here. In the proposed system, image keywords are extracted in fully-automated fashion by the deviced image recognition system. Image keywords used here is a collection of recognized objects in the image, where recognition levels are allowed to be intermediate or imperfect. In other words, each of these keywords has a various abstract level; some are very high and satisfactory, but other are very low and unsatisfactory. We introduce the concept of "recognition thesaurus" for managing these various level keywords successfully. "Recognition thesaurus" inherits the topology of the image recognition system, and consists of the relations between the different level of keywords. When someone wishes to retrieve images, he addresses a conceptual request to the system. The system interprets the request and makes an expanded keywords, then the system seeks images which satisfy the "expanded" keywords by using "recognition thesaurus". As an embodiment of this concept we implemented an image database with various types of sports scenes. Retrieval evaluations reveals the effectiveness of the proposed method.

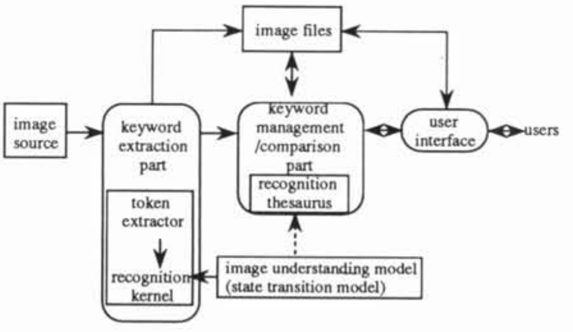


Figure 1: basic framework of proposed system

2 SYSTEM FRAMEWORK

Figure 1 shows the framework of the proposed system.

A keyword extraction part prepares image keywords from each image datum using an image recognition/understanding system. The image recognition/understanding system used here, makes a collection of recognized objects from an image. The most important feature of this recognition system is that achieved recognition level of each object can be accessed from outside even if it's an imperfect and intermediate one. We realize this feature by using our state transition model[5] as described in Section 4.1.

In the keyword management/comparison part, extracted image keywords are managed as a recognition thesaurus, detail of which is discussed in Section 3. Through this recognition thesaurus, image recognition mechanism is shared with the keyword extraction part. The user-interface analyzes users' requests and displays retrieved images.

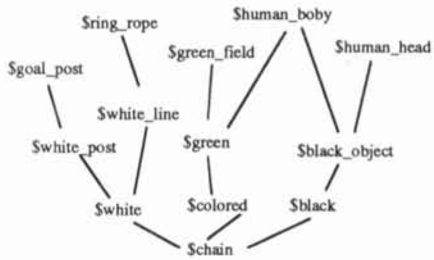


Figure 2: a typical state transition model

3 RECOGNITION THESAURUS(RT)

Here we describe "recognition thesaurus(RT)" which manages image keywords extracted with the image recognition system. RT is a kind of thesaurus. Ordinary thesauruses group keywords together according to their concepts or semantics, but RT groups according to their recognition level.

When we recognize images using a general purpose image recognition system such as a state-transition model[5], we often face the problem that some objects in an image cannot reach satisfactory recognition states because of failing to find recognition rules corresponding to the situations, on the other hands, in the ordinary database keywords of satisfactory levels of recognition are only used in the application or retrievals. Therefore, a complete image recognition system is required, and it is why fully-automated keyword extraction has not been realized.

To cope with this problem, we propose here the use of realized recognition levels, even if they are unsatisfactory ones. Figure 2 shows one typical example of the state transition model in our general purpose recognition system. The upper-most level corresponds to satisfactory levels and often gives conceptual keywords. Another recognized objects are unsatisfactory ones. Our idea is to employ these unsatisfactory recognition levels as keywords. To realize this idea, the recognition tree (i.e. Figure 2) is used for restoring the keywords. In that sense, we call this tree as "Recognition Thesaurus(RT)". For example, "$green" or "$black_object" in the tree of Figure 2 is used for a substitutional keyword of "$green_field" or "$human_head".

4 CONSTRUCTION OF SPORTS SCENE IMAGE DATABASE

As an embodiment of the general framework in Section 2 and 3, we implemented an image database with various kind of sport scenes in which automatically extracted keywords are managed by RT.

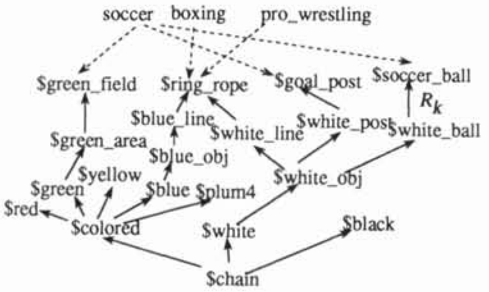
4.1 Keyword extraction

Firstly, let's discuss an image keyword extraction system in our system. As described before, a flexible image understanding system using a state-transition-type image understanding model[5] is employed for the purpose of both generality of the system and realization of RT described in Section 3. In this image understanding system, firstly, tokens, which are basic units for image understanding, are extracted from an image, and then, extracted tokens are labeled according to a state-transition-type recognition rules. Figure 3 shows one part of system description. A state transition model consists of a state transition graph and state transition rules. In Figure 3, transition rules are written in Prolog. Users can generate easily their own recognition system by giving these transition type recognition rules. Each token transfers its state according to its current state, its geometric parameter(color, shape, area), its environment situation, etc., until any token reaches successful state, or reaches transfer saturation.

We describe the example of the usage of the extraction and its relation with RT (refering to Figure 3). For example, as a result of recognition, segment (i) in the image (refering to Figure 4(b)) has reached $white_object state. This is actually a soccer ball, but its attributes and its environment circumstances do not allowed to satisfy the rule R_k for transition to $soccer_ball state. In such situation the system registers the segment as the state $white_object, and when images which include soccer_ball object are requested in the image retrieval process, the system searches not only data given the keyword "$soccer_ball" but also data given the keyword "$white_object". For this purpose, a set of the topological relation between all states(like the relation between the the state $white_object and state $soccer_ball) should be maintained in the retrieval system, which corresponds to the RT. In other words, the topology of the state-transition-type image understanding model itself can be used as "recognition thesaurus".

token extraction

In the image recognition system above, we need to extract tokens which is a basic unit of recognition from images. Because this image database is a full-color sports scene image database, segments extracted by a color segmentation system are selected as tokens. A color quantization method developed by our groups[6] is adopted as the color segmentation system, because it can realize fastest segmentation among various color segmentation including Orlander's method. Each token has its attributes like color informations (hue, saturation, intensity) of the segment, area, and shape. These attributes with their relations are used to construct the state transition rule for recognition.



(a) state transition graph

```
'$white_obj'(C, A, W, H, R) :-
'$get_corners'([_, _]),
W3 is W * 3,
H > W3,
A > 2000,
'$transform'('$white_post'(C, A, W, H, R)).
'$white_obj(C, A, W, H, R) :-
C > 0.8,
A > 2000
'$transform'('$soccer_ball'(C, A, W, H, R)).
```

(b) state transition rules

Figure 3: a part of a state transition graph and transition rules

state transition model

We construct a state transition model for labeling extracted tokens(, whose one part is given in Figure 3). As describe above, the state transition graph can be used for "recognition thesaurus" of this database.

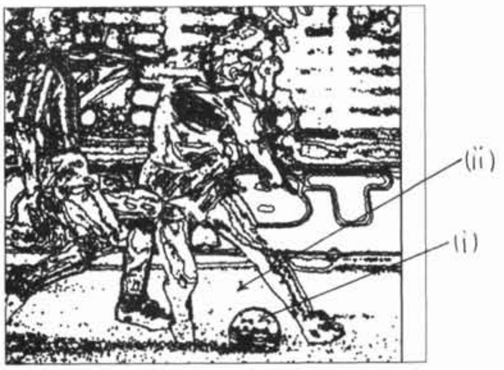
Figure 4 shows an example of these image keywords extraction. Figure 4(a) with the chain codes of all segment is displayed in Figure 4(b). All segments are labeled with the state-transition-type image recognition model. For example, the segment (ii) in Figure 4 are labeled "green_field". All the segments(except too small region) are registered in the database with "recognition thesaurus"(Figure 3(a)).

keyword management

As describe before, in the keyword management in our database not only satisfactory recognized results corresponding to conceptual level, but also unsatisfactory results corresponding to intermediate levels of the recognition tree should be maintained with this topology.

Figure 5 shows the image(i.e. token) data format for this management. Each object possesses its objectID, imageID, keywordID, and attributes. These object data are also able to be thrown into the existing multi-dimensional database system to correspond to the retrievals including the spatial relations between

(a) original image



(b) edge image

Figure 4: example of image keyword extraction

obj ID	img ID	keyword	attributes						
			area	circ.	color			location	
					hue	satu.	inten.	x	y
1	1	$green_field	41458	0.097	2.43	16.86	75.12	240	215
2	1	$white_obj	2409	0.289	2.92	3.63	97.49	415	129
3	1	$white_ball	1046	0.332	2.92	3.63	97.49	303	197
4	1	$black_obj	792	0.130	2.88	7.00	54.61	126	64
5	1	$black_obj	1025	0.029	2.88	7.00	54.61	215	115
6	1	$black_obj	4288	0.030	2.88	7.00	54.61	302	203
7	1	$black_obj	3248	0.053	2.88	7.00	54.61	184	203
8	1	$black_obj	657	0.256	2.88	7.00	54.61	28	156
9	1	$blue_obj	533	0.156	4.23	10.24	67.75	342	71
10	1	$blue_obj	528	0.024	4.23	10.24	67.75	134	78

Figure 5: data format of object data

objects. By this functions, such retrievals as "search images including 'human' in the upper-right hand part, can be easily realized. As a spatial management system, the 8-dimensional GBDtree[4] is used.

4.2 Retrieval process

Users(retrievers) are supposed to want to retrieve images by their conceptional requests, which includes various type retrievals concerning to not only contents of the objects in the image, but also scene contents, exciteness, etc. As describe before, in our system, managed keywords are based on the objects in the image with RT, so abstraction level of most are not necessarily conceptual and multi-view. Therefore, the most important function is that the system must interpret the requests and convert them to the objective level

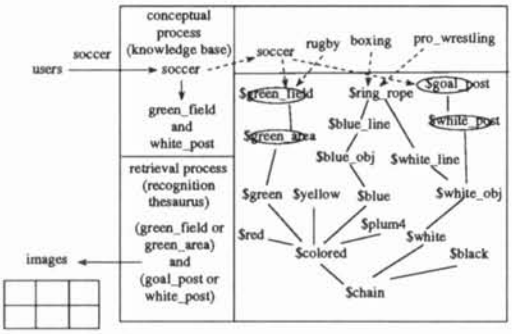


Figure 6: retrieval flow

description.

In our system the relation of the requests and the objective-level keywords is stored like a knowledge base. In the present embodiment, this relation is represented as linking of the request and the objects to be included in the required images. This method is illustrated in Figure 6. For example, if user gives a conceptual level request including "soccer", the system convert it to lower level(i.e. objective-level) description "$green_field and $goal_post" using this knowledge base. To realize the knowledge base in flexibility and gradual fashion, interactive indication of some concept is allowed.

4.3 Experimental results

To evaluate the feasibility and effectiveness, an experimental system based on the above idea has been implemented on our workstation(Sun 3) in language C. System size of basic part is about 2000 lines. In the experiment, 50 scene images from 10 kinds of sports, including soccer, pro-wrestling, track-and-field, voleyball, etc. Various kind of retrieve experiments have been tried. Results are fairly satisfactory except that in most of retrievals some "noise" images reflecting the RT structure with lower-level objects description have been mixed in the desired images.

Figure 7 shows one example of such retrievals. A user requested to retrieve "soccer scene". Firstly, the system search the keyword "soccer". But it failed because such conceptual level keyword could not be extracted automatically. Then, the system converted the requests to lower concept "green_field, goal_post" "soccer_ball", and "stadium_fence" in the RT, and each of objects including the keywords ("green_field", "goal_post", etc.) are retrieved from the object database. The images with higher certainty defined by low objects which include all of the required objects are displayed(as can be seen in Figure 7).

Seeing the displayed images, the required "soccer_scene" images are included successfully, but unnecessary scenes are also included as noise in this case, because these scene include similar necessary objects of lower abstraction level.

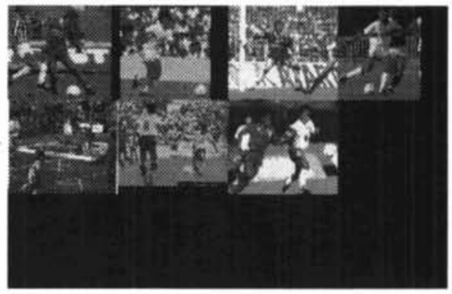


Figure 7: retrieval result for request "soccer scene"

5 CONCLUSIONS

A new image database system with capability of fully-automated image keyword extraction, using "recognition thesaurus" has been presented. "Recognition thesaurus" plays an important role to realize the practical combination of difficult image understanding and relatively satisfactory retrievals. In other words, we could say that this system can realize retrieval with fully-automated keyword extraction at about 70~80% using image understanding technique at about 50%.

Hereafter, we intend to increase example images more powerful and user-interface more flexible and improve the framework including the image understanding function.

REFERENCES

[1] M.sakauchi, Y.Ohsawa: Image Database Technology, Journal of the Institute of Electrial Engineering of Japan, Vol.105, No.5(1985).

[2] M.Sakauchi: Image Retrieval Technology, Journal of Institute of Electronics Information and Communication Engineers, Vol.71, No.9, pp.911-914(1988).

[3] Special Selection on Image Databases, IEEE Transaction on Software Engineering, Vol.14, No.5, pp.608-688(1988).

[4] Y.Ohsawa, M.Sakauchi:A New Tree Type Data Structure with Homogeneous Nodes Suitable for a very Large Spatial Database, *Proc of IEEE Sixth International Conference on Data Engeneering*, Los Angels, 13-1(1990).

[5] S.Satoh, Y.Ohsawa, M.Sakauchi:Drawing image understanding framework using state transition models, *10th ICPR*, pp.491-495, IEEE(1990).

[6] Y.H.Gong, Y.Toriumi, Y.Ohsawa, M.Sakauchi: An Interactive color Image Quantization Method with Flexibe and Subjective Color Selection Capability, The Journal of the Institute of Television Engineers of Japan, Vol.45, No.1, pp.86-93(1991).