Fifth IFSA World Congress (1993), 1309-1312

An Image Retrieval System with Adjustment for Human Subjectivity

Shigenobu FUKUSHIMA and Anca RALESCU [1,2]

Laboratory for International Fuzzy Engineering Research
Siber Hegner Bldg. 4Fl., 89-1, Yamashita-cho, Naka-ku, Yokohama-shi, 231, JAPAN

Abstract

We present a flexible retrieval system of face photographs based on their linguistic descriptions in terms of fuzzy predicates. While natural for describing a face, linguistic expressions are also subjective, which affects the retrieval result. Thus, the capability of a retrieval system to adjust to different users becomes very important. In this research we use fuzzy logic techniques, for describing image data, inference for retrieval and adjustment to a new user. Experimental results of the adjustment are also included.

key words: fuzzy sets, matching, linguistic queries, flexible retrieval

1. Introduction

In recent years, human friendly systems, their design and implementation, have become a central issue in the study of intelligent systems. A minimum requirement for such systems should be the acceptance by the system of a natural language like input from the user. In this paper we consider the problem of image retrieval based on linguistic queries. Thus we are faced with two issues: linguistic modeling and image retrieval. Research in knowledge acquisition, and knowledge representation [4] focuses on deriving theories, more or less general, for natural language representation and understanding. While interesting and appealing from theoretical point of view, these theories are usually difficult to implement. Other work coming from the field of fuzzy engineering [5], deals with linguistic modeling of numeric data, using fuzzy sets. Integration of ideas from the two fields has been proposed and explored in [3]. On the other hand, a robust method of image retrieval with application to face recognition has been developed in [6] without any reference to the linguistic description of the image.

A different, more applied point of view was adopted in [2] to investigate fuzzy logic based linguistic modeling and retrieval of image data. The system, shown in Fig.1, produces linguistic descriptions of human face photographs. The goal of the system is to obtain such descriptions which agree with the impression made on a user by the photograph in question such that the description can be used as query. While natural language-like descriptions of image data are very natural, their reliability is somewhat fragile. We can say that in some sense such descriptions are neither complete, nor sound. Furthermore, it is very difficult to guarantee that a linguistic description is correct in the sense of being entirely accepted by all users. This is due mainly to the highly subjective nature of such descriptions, which vary among individuals or groups and is influenced by the format of questionnaires. The work presented in this paper builds on previous work [2] implemented in a image modeling and retrieval system and the goal is to present the results obtained from adding an adjustment capability to this system.

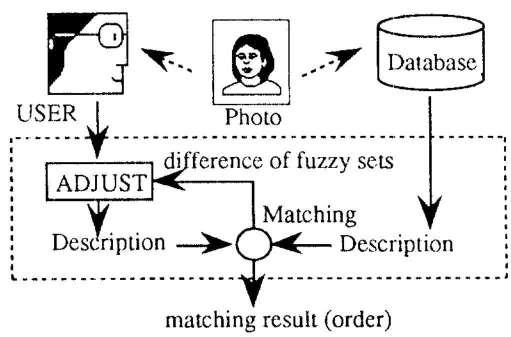

Fig. 1 Retrieval system with adjustment

2. Comparing two models of the same photograph

To increase the flexibility of the linguistic retrieval we will endow our system with an adjustment capability. We take as point of departure for our approach the following observation of the human behavior: individuals from different cultures, environment, etc. are able to communicate only to the extent that they are able to (temporarily) adjust to each other. In this section we will discuss the general idea behind adjustment as well as

[1] This research was partially supported by the NSF Grant INT91-08632

[2] On leave from the Computer Science Department, University of Cincinnati, U.S.A.

detailed situations in which an adjustment (and its type) is needed.

2.1 The general idea of adjustment.

The idea is that an adjustment of the system to a particular user should improve the retrieval. This adjustment, as opposed to learning, is temporary, that is, it is 'remembered' by the system only for one user at a time. The adjustment is based on recording and analyzing the history of the user's input. The essential point is to observe past behavior of a user to derive some global knowledge of this user's tendencies and use them in future queries provided by the same user. The retrieval system we have in mind will have an additional block, responsible for adjustment, as indicated in Fig.1 The difficulty of this system lies in its implicit nature: we avoid asking a user great effort, thus the information obtained may not be enough for understanding the user's view point. Therefore, we detect and treat only the user's tendencies.

2.2 Subjectivity of users

In order to understand what happens when a retrieval operation fails, we analyzed the results obtained from linguistic queries coming from several users. We found that tendencies of behavior can be identified as follows:

(a) overstatement / understatement

This corresponds to the user tendency to give extreme values when describing the characteristics of the photograph. For example, in case of overstatement the system description 'somewhat big/somewhat small' is expressed by the user as 'big/small'. This tendency appears globally, that is, across all(most) features.

(b) shift in one direction

This corresponds to the case when there is a consistent translation-like discrepancy between the system descriptions and the user. For example, this appears when the system's description 'somewhat big' is expressed 'big' and 'small' is expressed 'somewhat small'. If we imagine a scale on which 'small' and 'big' are represented from left to right, the previous example corresponds to a translation to the right. The cause for the shift discrepancy seems to originate from different meanings of 'middle'/'medium', among different users (and system as well). This tendency appears locally (for some features).

(c) randomness

We assume that the discrepancies between the user and system which cause failure in retrieval, but for which tendencies (a) or (b) cannot be identified, are due to random assignment of the features to the description. The causes for randomness are not very clear.

2.3 Matching of fuzzy sets

Since the features of photographs are expressed as fuzzy sets, the basic mechanism of the retrieval operation is the match of fuzzy sets. In this paper the match operation proposed in [1] is used:

$$\mathrm{Match}(A,B) = O(A,B)\ NF(A)\ NF(B)$$

where O(A,B) is a measure of overlap of A and B, and NF(A) is a measure of fuzziness of A. This measure is defined with respect to the concept of "fuzziest" fuzzy set. The choice of this set is, to a large extent, a matter of convention. It is usually made such that it agrees with some intuitive meaning of the notion of fuzzier/fuzziest. For example, according to [7], [8] the property of a fuzzy set of being fuzzy is best captured by the inability to distinguish between the set and its complement: the less we can distinguish between these the fuzzier the set is. Moreover this property must take place globally. Thus, the fuzzy set with constant membership function $\mu(x) = 0.5$ is the fuzziest fuzzy set.

In this paper we represent each linguistic expression as a triangular normal fuzzy set. For these sets we defined the fuzziest fuzzy set to be the triangular set whose support is equal to the universe of discourse. Let A and B be two triangular fuzzy sets on the universe of discourse [a, b], with supports S_A, S_B respectively. Let A_{AB} denote the area of the intersection of A and B. Then we define the overlap of A and B as

$$O(A, B) = \frac{S_A + S_B}{S_A S_B} A_{AB}$$

It follows from [7] that for a triangular fuzzy set A its fuzziness F(A) is defined as

$$F(A) = \frac{S_A}{b-a}$$

and its non-fuzziness, NF(A) is given by

$$NF(A) = 1 - F(A) = 1 - \frac{S_A}{b-a}$$

Thus the match of A and B is then:

$$\mathrm{Match}\ (A,B) = \frac{S_A + S_B}{S_A S_B} A_{AB}\left(1-\frac{S_A}{b-a}\right)\left(1-\frac{S_B}{b-a}\right)$$

This matching method considers not only the amount of agreement between the fuzzy sets, but also how fuzzy the two fuzzy sets are. More specifically, if the two fuzzy sets are identical (or alternatively if matching a fuzzy set to itself), the result of their perfect overlap is affected by their fuzziness: the fuzzier they are (that is the larger the support) the smaller the final matching result will be. Thus the fuzziness affects the confidence in these fuzzy sets. We found this matching to be especially adequate for the adjustment of the retrieval system: When an expression appears to be not trustworthy, we can increase the support of its associated fuzzy set, making it fuzzier, decreasing the matching degree, and hence its impact on the retrieval.

3. Adjustment of the user's model

3.1 Recording the user-system relation

The adjustment is based on recording past performance of the system for a given user. In order to classify the behavior of a given user we introduce the following notations to record overstatement/understatement:

'+' : overstatement
'-' : understatement
'=' : agreement (this corresponds to the case in which the user and system model are either identical, or within a previously defined tolerance ε).
'?' : not '+' or '-' or '=' (The adjustment of the system in this case will be the same as in the overstatement case.)

For adjustment of shift and randomness we consider the quantity (user input - database) between the centers of gravity of the system's descriptions and the user's input. We distinguish between possible values of this quantity as follows:

'−3': (user input - database) $\leq$ -d3
'−2': -d3 < (user input - database) $\leq$ -d2
'−1': -d2 < (user input - database) $\leq$ -d1
'0' : -d1 < (user input - database) < d1
'+1': d1 $\leq$ (user input - database) < d2
'+2': d2 $\leq$ (user input - database) < d3
'+3': d3 $\leq$ (user input - database)

where, d1, d2, d3 are some constants chosen at the design phase of the adjustment system. In the current implementation the values d1=0.5, d2=1.0, and d3=1.5 were used. In this connection, the center of gravity of the user descriptions can be 1, 2, 3, 4, 5 and that of the system descriptions can be a real number from 1 to 5.

3.2. Adjustment for overstatement and understatement

The adjustment procedure is as follows:

1. Collecting data about input: For each input photograph i (6 descriptions for each photograph), determine D(i,+) and D(i,-) where D(i,+), D(i,-) is the number of descriptions of '+' or '?', '-' respectively.

2. Collecting data across inputs: Concerning of the history of input, count T_+, T_- where
T_+: number of times when D(i,+) > D(i,-)
T_- : number of times when D(i,+) < D(i,-)

3. Calculate $T_{+-} = T_+ - T_-$

4. For each feature, determine E_+, E_-, $E_=$, the number of '+', '-', '=' respectively by counting.

5. Determine the quantity R_{+-} as follows:

$$E_{+-} = \begin{cases} \dfrac{E_+}{(E_+ + E_- + E_=)} & \text{if } T_{+-} > 0 \\ 0 & \text{if } T_{+-} = 0 \\ \dfrac{E_-}{(E_+ + E_- + E_=)} & \text{if } T_{+-} < 0 \end{cases}$$

6. Determine the index of adjustment (I+-): Let FR(_, _) denote a fuzzy reasoning scheme, decided experimentally. Then I_+=FR(P_{+-}, R_{+-}). The index of adjustment is used to decide the amount of adjustment

as in step 7 below:

7. Support modification: Increase the support of the fuzzy set associated to an expression by the amount D_{+-} =$(D_{ud})(I_{+-})$. (D_{ud} denotes the distance between the centers of gravity of user's input and the system's 'middle').

The resulting adjustment is shown in Fig.2(a), Fig.2(b) for the cases of overstatement and understatement respectively.

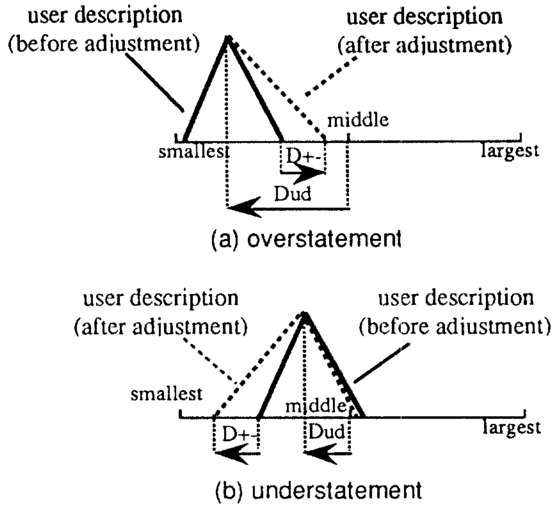

Fig.2 Adjustment for overstatement/understatement

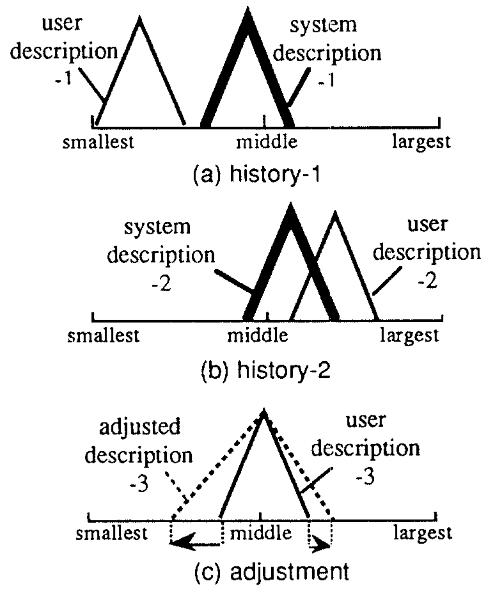

Fig.3 Adjustment for shift and randomnes

3.3 Adjustment for shift and randomness

The adjustment for shift and randomness is done locally for individual features. Since the difference between 'shift' and 'randomness' is not very clear, we use a simple but robust method as follows. The essential idea is to increase the support of every fuzzy set from the user's input in accordance with the accumulated difference between the user input and the database. However, instead of accumulating the differences precisely, we classify the differences into 7 levels, count the frequency within each level and calculate the amount of support increase based on these frequencies. The adjustment procedure is then as follows:

1. For each description, count the number of all relations , '-3', '-2', '−1', '0', '+1', '+2', '+3' and their respective sums, N-3, N-2, N-1, N0, N+1, N+2, N+3.

2. Let Db_l, Db_r denote the amount of increasing the support of a fuzzy set to the left and right respectively. These amounts are given by the following equations:

$$Db\_l = 0 \vee (N\text{-}1 + 2\,N\text{-}2 + 3\,N\text{-}3 - N0)\, d0$$

$$Db\_r = 0 \vee (N\text{+}1 + 2\,N\text{+}2 + 3\,N\text{+}3 - N0)\, d0$$

where, $\vee$ denotes the maximum operation, and d0 is a constant determined experimentally. In the current implementation d0 = 0.1.

3. Repeat 1 and 2 for all the descriptions input by the user.

Fig.3 shows this adjustment on the basis of a history of two previous inputs for a feature: history-1, history-2.

3.4 Global adjustment (Combination of two adjustments)

Discrepancies due to overstatement/understatement on one hand, and shift or randomness on the other hand, can appear at the same time. The previous adjustment procedures may prescribe conflicting actions regarding the increase of supports. To combine these we use an evidential reasoning like paradigm based on the amounts of support increase prescribed (to the left or to the right). Let Ix, denote the importance of support increase, where x=a if the adjustment is due to overstatement/understatement, and x=b if the adjustment is due to shift or randomness. Ix is defined, in terms of the quantities Dx_l, Dx_r as follows:

$$Ix = (Dx\_l + Dx\_r) / (Da\_l + Da\_r + Db\_l + Da\_r)$$

Then the combined adjustment is determined by Dab_l, Dab_r which are given by the following equations.

$$Dab\_l = Da\_l\ Ia + Db\_l\ Ib$$
$$Dab\_r = Da\_r\ Ia + Db\_r\ Ib.$$

That is, the combined adjustment is a convex combination of the single adjustments weighted proportionally.

4. Simulation results

To evaluate the adjustment procedures described above the following simulations were performed for 18 input photographs (2nd to 19th input):

(1) retrieval without adjustment

(2) retrieval with adjustment for (i) overstatement /understatement

(3) retrieval with adjustment for (ii) shift and randomness

(4) retrieval with adjustment for both (i) and (ii)

All the adjustments use at most 10 input. Note that all the parameters for the adjustment were decided experimentally based on the simulation results for five subjects and the same parameters were used for the simulation of the remaining subjects.

The results shown in Table 1 show the ranking average in the retrieval of the 18 input photographs, and the improved ratio compared with the retrieval without adjustment.

In average, we obtained 4.65%, 3.80%, and 4.81% improvement for the adjustment for (i), (ii), and (i)(ii) respectively. Table 2 shows the effect of the adjustment in terms of the number of cases in which adjustment brings better order, worse order and the same order respectively. From this table it follows that the combination of two adjustments results in a clear improvement of the order: 163 (or 38%) of the cases were ranked better after the combined adjustment, while for 199 (or 41%) the order remained unchanged. Only 81 (21%) cases resulted in a lower rank than without adjustment. This analysis points to the fact that it may not be possible to realize an absolute improvement, but the exact meaning of this, as well as alternative evaluation procedures must be studied further.

	no adjustment	adjustment (i)	adjustment (ii)	adjustment (i) (ii)
average of order	5.566	5.327	5.355	5.311
ratio of improvement		4.65%	3.80%	4.81%

Table 1 result of adjustment (average of order and improvement)

	adjustment (i)	adjustment (ii)	adjustment (i) (ii)
better order	145	140	163
worse order	103	90	91
same order	181	199	175

Table 2 result of adjustment (better order and worse order)

5. Conclusion

The purpose of this study is to treat the human subjectivity as it appears when humans describe images. The application considered focused on the 2-d images of human faces. However, the work presented here can be applied to the description of any images (such as natural scenes for example) where human subjectivity plays an important role. From our analysis of human behavior we hypothesized three causes for discrepancies between descriptions of the same image given by different human subjects. These were tendencies for overstatement/ understatement, shift and randomness. An important aspect of our study is the use of fuzzy logic: for expressing attribute values, their matching and for reasoning. Future work along these lines should consider obtaining an adjustment method suitable to an individual user and which takes into account more of the history of that user.

References

1. Gasos J., Fukushima S., Ralescu A. "Fuzziness Dependent Matching of Fuzzy Sets" (in preparation)

2. Nakayama M., Norita T., Ralescu A. "A Fuzzy Logic based Qualitative Modeling of Image Data" Proceedings of IPMU'92, pp. 615-618

3. Ralescu A., Narazaki H. "Integrating Artificial Intelligence Techniques in Linguistic Modeling from numerical data", Proceedings of IFES'91, pp. 328-337.

4. Sowa J. F., (eds.) (1992) *Principles of semantic Networks*. San Mateo, CA: Morgan Kaufman Publishers.

5. Sugeno M., Yasukawa T. "Fuzzy Logic based Qualitative Modeling" IEEE Transactions on Fuzzy Systems (to appear in January 1993).

6. Turk M. A., Pentland A. P. "Face Recognition using Eigenfaces", Proceedings of CVPR'91 pp 586-591.

7. Yager R. R. "On the measure of fuzziness and negation. Part I: membership in the unit interval." International J. of Man-Machine Studies, 11, pp. 189-200

8. Yager R. R. "On the measure of fuzziness and negation. Part II: Lattices", Information and Control 44, pp. 236-260.