Towards Mutual Comprehension through Interaction

Nadia BIANCHI BERTHOUZE
*Intelligent Systems Division,
Electrotechnical Laboratories
Tsukuba, Japan
bianchi@etl.go.jp

Paolo BOTTONI
Department of Computer Science
University of Rome
Roma, Italy
bottoni@dsi.uniroma1.it

Toshikazu KATO*
Department of Industrial
Engineering
Chuo University
Tokyo, Japan
kato@etl.go.jp

ABSTRACT

We explore interaction as a basis for human-computer comprehension. Perceptual experience is organised through categories establishing the kind of distinctions imposable on perceived phenomena. An interactive tool exploiting a similar categorisation in graphics is integrated into a system for subjective retrieval of images, so that the user interacts with active regions supporting the same type of distinction.

Keywords

Lexical categories, active region, Kansei interaction.

1. INTRODUCTION

Personalisation is fundamental to structure new computer-based activities. Personal assistants could learn owners' subjective preferences in complex cognitive processes such as picture appreciation, or expression of aesthetic impressions.

Natural language is the tool of choice for communication among humans, but its use as a tool for human-computer communication is more problematic [6], and generally restricted to specific, often technical, fields. On the other hand, direct manipulation has become ubiquitous in human-computer interaction (HCI). We exploit it as a basis for a com-munication language between a personal assistant and its owner, so that they can come to agree on the meaning to be ascribed to terms, even in case of subjective impressions.

The work capitalises on an architecture for learning subjective impressions [4] and on the notion of *active region* from the visual language and HCI communities [7, 11: interaction is a means to build a common lexicon and a sharable organisation of perception [3]. A tool for construction of term meaning is integrated into MIKE, an environment for retrieval of multimedia material based on subjective impression. Here, we focus on the human ability to use terms to refer to distinctions in the perceived phenomena. In expressing distinctions, perceptual adequacy is separated from linguistic competence, and the latter from conceptualisation capability. The application domain internal structure is used where available.

2. BACKGROUND

The problem of defining a common language between user and computer is usually solved by recurring to a limited vocabulary, or embedding natural language abilities in the system, possibly integrated with forms of direct manipulation [1]. However, lexicon is not a stable entity, and even private lexicons within a community can get modified [15].

When interacting on subjective impressions, the meaning of terms may differ among subjects, and is hardly definable via a linear mapping from physical features. ARTMUSEUM uses a fixed lexicon and users can assign grades to each term, in order to assess a picture from a database [14]. In its recent evolution, new terms, possibly based on existing ones, can be associated with nodes in the upper layer of a neural network whose input nodes are colour features measured in the image [2].

Colour features are popular in image retrieval systems, since colour is easily segmented, relies on one of the most salient aspects of an image, and usually carries sufficient semantic content to support forms of content-based retrieval [8, 13, 9]. However, impressions can derive specific arrangements of coloured regions, as well as from measurable features, such as size, shape, texture, etc. Moreover, not all types of impression derive from the same features. A system must learn not only which measured values correspond to a certain term, but also which features are relevant for a specific judgement.

We start from a categorisation of adjectives for properties of an image or of regions therein. Overall impressions are conveyed by *observation* adjectives, not corresponding to single measures or their specific combinations. They are associated with configurations of a neural net and learned by the system in a supervised way [2]. Users can *externalise* the way in which the judgement is formed, by pointing at some features of the image. Adjectives usually refer to *size and shape* properties, and are associated with sets of values in the possible outcomes of single measures. We propose to use suitable forms of visualisation to help users organise their externalisation, so that a common organisation of users' perceptions and system's measures can be achieved. The visualisation captures some form of isomorphism between categorisation of perceptions and of graphical structures. The relation between orders in the domain and orders perceived in individual graphical structures is studied in [10]. Properties of graphical structures supporting correct visualisations are studied in [16]. These studies have taken into account static aspects of graphical structures. Dynamic aspects, considering orders in visual interaction, have been studied in [5], where focus was on transformations of graphical structures. Here, we exploit very

simple forms of visualisations, but propose a general framework in which different visualisations can be developed.

3. SUBJECTIVE IMPRESSIONS

Verbalisation of cognitive or perceptual phenomena is mediated by speakers' culture and experience, in particular their linguistic competence. The relation between objective properties and subjective responses is difficult to make explicit and is typically variable. In our framework, it emerges from two processes: (1) a decomposition of subjective impressions into collections of sub-impressions and low-level features, and (2) a synthesis of user's intervention, available physical observations and past experiences. On the other hand, terms not referring to immediate perceptual experience depend on the ability to organise one's perceptions. While judgements which refer to perceptual aspects are often graded, emotional or cultural judgements seem to have a rather dichotomic nature.

However, the ability to draw distinctions at a perceptual level does not imply the use of a specific term for each distinction. Also, the ability to organise perceptions into a concept does not correspond to that of defining an explicit term to denote it, or of using the term in the sense accepted by other speakers. Finally, rather than on the correct usage of terms, it is important to agree on the semantic area covered by the term in the shared lexicon. This is facilitated if one can refer to the structure of the underlying domain. In particular, graded judgements are expressed w..r.t a linearly ordered model, while emotional judgements are predicates assessing whether something corresponds or not to a prototype.

4. MAPPING TO STRUCTURES.

We propose a tool by which users manipulate meanings and terms to establish a private, expandable lexicon for com-munication with the system, and develop forms of conceptualisation and organisation of perceptual experience. We identify the following categorisations of perception.

- **opposition**: dichotomies corresponding to some property, or to its opposite, or lack thereof.
- **linear ordering**: perceptions on a refinable scale from a minimum to a maximum;
- **abstraction:** scale of abstraction;
- **distribution**: things are simply perceived as "being there" and being distinguished from others.

Some graphical representations naturally support isomorphic organisations of the domains they represent. Thus, opposition is expressed via *inside/outside* relations; linear ordering is represented by a one-dimensional structure, with a reading convention; abstraction induces a partial order, represented by levels with the simultaneous use of two dimensions; distribution is represented by identifying arbitrary regions.

4.1 Communicating perceptual experience

With the tool, users define perceptual judgements by directly manipulating graphical structures, establishing the lexicon to communicate thier experience, get informed of the current usage of this lexicon within the system. In particular, graded judgements are expressed via interaction with a solid line (Figure 1). By clicking on a position on this line, the user creates a distinction, subsequently associating a set of adjectives with it. At any moment the user can modify the distinctions, modify the labels, associate a new label to a distinction previously left anonymous, introduce new distinctions or eliminate old ones. Similarly, assertive categories relative to dichotomic perceptions (e.g. present / absent) are judged with respect to dichotomic terms (Figure 2).

Such an organisation of the perceptual experience with respect to the lexicon can be performed for the features measurable by the system. Distinctions can be identified also without reference to specific pictures, as well as on specific zones of images. A new learning phase need occur only when distinctions, not terms, are modified.

4.2 Communicating emotional experience

A "flat" organisation of the space of terms does not discern between terms evoked by direct stimuli (e.g. a picture may be *bright* in that it presents vivid colours) and those corresponding to mediated impressions or complex associations (e.g. a picture may be *appeasing* in that it has an equilibrium in the distribution of soft colours). On the other hand, the same picture may be evaluated in different ways according to different organisations of a same perceptual experience: an appeasing picture may also be *romantic*, in a different categorisation (thematic rather than emotional). On the contrary, a *romantic* picture may also be *disquieting*, if it evokes romantic emotions by violent clashing of colours. These high level experiences can be managed by letting the user interact at different levels with the system.

On the one hand, we see emotional judgements as dichotomic ones, and let the user ask for pictures which present or lack some qualities. On the other hand, we help the user relate these emotions to its constituents. The system can thus recognise that images a user has deemed as *appeasing* have, say, "*large* areas in *pastel* colours AND areas are prevalently extended in the *horizontal* direction". The user can intervene to present other examples of *appeasing* pictures with different aspects, accept the proposal of the system, or edit the definition. Moreover, users could indicate that only some of the features indicated by the system have to be taken into consideration in formulating such semantics, or that different features, among those available to the system, must be considered.

5. DISTINCTIONS AND TERMS

We propose a Java open implementation where new terms, categories and features are added during system usage. The linguistic notions of *category* and *distinction* are accessed and manipulated by interacting with *active regions* and *active sectors*. Categories are paired with classes of features supporting different types of assessment or measure, and sets of terms are associated with the distinctions they designate.

A `Feature` is an attribute of a multimedia entity that supports setting and getting an `Object` value; a `Category` sets a partition of a `Feature` domain, composed of `Distinction`s, which are aggregations of `Separator`s. Terms are associated with a `Distinction` and obtained from a `DataSource`, e.g. a thesaurus, or `TextField`. Different types of `Category` can be devised, such as `GradedCategory` for `Feature`s whose domain supports a linear order, or `AssertiveCategory` for judgements in the form of dichotomies. An `ActiveRegion` represents a `Category`. Users interact with its `ActiveSector`s (representing the `Distinction`s) and redefine the `ActiveSeparation`s (representing the

Separators) which uniquely define them. According to the type of Category, Regions, Sectors and Separators have different graphical types.

Figures 1 and 2 both present on the left an ActiveCanvas containing the ActiveRegion related to the explored Category, namely an ActiveLine in Figure 1, and an ActiveEnclosure in Figure 2. In Figure 1, the user is defining the distinctions and the associated terms for the category *Aspect*. In Figure 2, the user is defining a collection of terms to be used for assessing some feature in the image. The term *positive* is associated with the *inside* distinction.

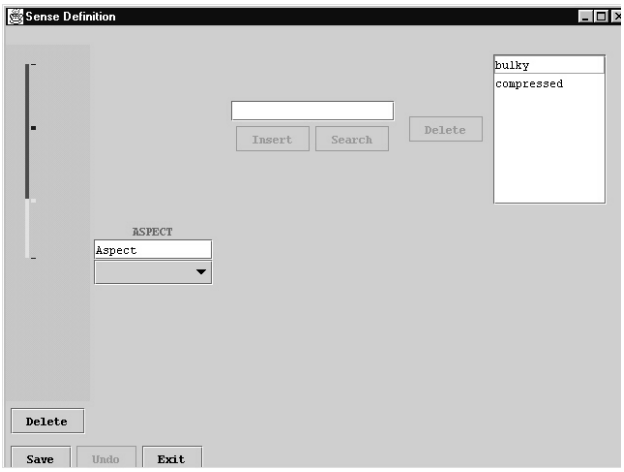

Figure 1. The frame for management of graded categories. The user is defining distinctions for the *Aspect* category.

In both cases, users introduce new terms in the DataSource area on the right, associated with the current Distinction, represented by the current ActiveSector, eliminate terms from the current distinction, or query the system to find whether a given term has been introduced. The ActiveRegion is presented in red, the ActiveSector in yellow, and the ActiveSeparators in blue. Consistency is guaranteed at any time of the visible set of terms, the current Distinction, and the current ActiveSector in ActiveRegion.

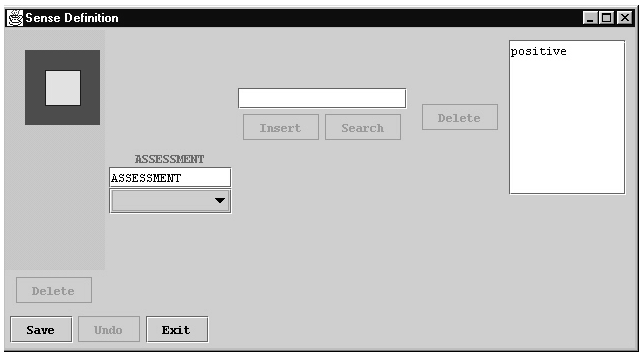

Figure 2. The frame for management of assertive categories. The user is adding terms to the default assertive category.

A term can be transferred from a distinction to another in two ways: 1) delete a term from the list, modify the active Distinction, and reintroduce it; 2) double-click on the term in the list; when the term appears close to the ActiveRegion grab it and move it to a new distinction. Actions are enabled only in a correct state, e.g., separator removal is allowed only if at least two distinctions remain: the last ActiveLine separator cannot be removed, nor the only ActiveEnclosure separato. Such policies are managed by associating objects with events which may modify their enabled state.

6. A SCENARIO OF USE

Nathalie and the system (called Mr. Feel) are building a common lexicon for analysis of pictures based on subjective impression. They have agreed on the meaning of some basic terms, such as *big* or *dark*, and are trying to establish meanings for more complex terms, such as *happy*. Based, on what it already knows, Mr. Feel takes the initiative to look for *happy* images in its picture data base. In Figure 3 it proposes an example of *happy* picture; Nathalie counteracts with the introduction of the new term *scary*.

Now Nathalie explains the meaning of *scary* by using terms already in the lexicon, such as *big* and *red*, and indicating the physical features that gave rise to the judgement of *scary*. The regions in the right-most window are detected automatically on the basis of the chosen colour (red, green). The white region on the central window is drawn by Nathalie, who introduces the new term *elongated*. Rather than trying to explain *elongated* based on existing terms, or to let the system infer a new node in the network starting from this example, Nathalie chooses to directly teach the system her meaning for the term.

Hence, she loads the category *Aspect* in the categorisation environment to *Aspect* is relative to the feature computed as the ratio of height to width for a region. A linear order exists for it, and it is represented as a straight line. No terms have been defined for it yet, and only the two defaults distinction for a graded category exist. Nathalie decides to define a basic lexicon for this category and sets three distinctions, to which she associates the terms *elongated*, *balanced* and *compressed* respectively. She also introduces the term *oblong* for the first distinction and the term *bulky* for the third one.(Figure 1).

Once completed this definition, she goes back to the dialogue, where she can use the new terms. She can then refer collectively to the regions of *elongated* form and say that they *are arranged as "Prison-bars"*. The system recognises the construct *arranged as* and tries to extract the spatial relations among the indicated regions. Hence, it can further understand that a set of elongated regions look like *prison-bars* if they are parallel and close to each other, and that such an arrangement contributes to making an image *scary*.

7. CONCLUSIONS

The problem of building a common language between a user and a personal assistant is faced by exploiting the subjective ability of the user to provide different terms for the categorisation the user imposes on his or her perceptions. An interactive tool exploits an isomorphism between categorisations of graphical types and of structured domains. Users can associate terms with distinctions, edit the portion of the active region pertaining to a distinction, retrieve the distinction associated with a term, modify current assoc-iations, refine or merge distinctions, etc. Each action is

relative to the currently active sector, or makes a different sector active.

The tool is currently used offline in MIKE, a system for retrieval of multimedia documents by subjective feelings [4]. The architecture for complete integration has been designed. Besides subjective retrieval, applications are foreseen for document annotation, since the notion of active region can support several types of semantic definitions. The study of the isomorphisms between application domains and graphical structures is in general relevant in several HCI field.

8. ACKNOWLEDGMENTS

Partially supported by STA and MITI, Japan, and the EC. P. Bottoni was granted an STA fellowship at ETL.

9. REFERENCES

[1] A. De Angeli, W. Gerbino, G. Cassano, D. Petrelli, “Visual Display, Pointing and Natural Language”, *Proc. AVI ’98*, 164-173, 1998.

[2] N. Bianchi-Berthouze, L. Berthouze, T. Kato, "Toward an integration of subjective parameters in database browsing". In Y.Kambayashi *et al* eds., *Advanced Database Systems for Integration of Media and User Environments*, World Scientific, 227-232, 1998.

[3] N. Bianchi-Berthouze, L. Berthouze, T. Kato, "Understanding Subjectivity: An Interactionist View", *User Modeling 99*, 3-12, 1999.

[4] N. Bianchi-Berthouze, T. Kato, "An Interactive Visual Environment to Understand, Model and Exploit User Subjectivity: in Image Retrieving Systems", *ICIAP 99*, 981-986, 1999.

[5] P.Bottoni, M.F.Costabile, S.Levialdi, P.Mussio, "On orders in visual interaction", *JVLC*, 10, 343-380,1999.

[6] S.E. Brennan, “Conversation with and through Computers”, *User Modeling and User-Adapted interaction*, 1, 67-86, 1991.

[7] S.-K. Chang, T.Y. Hou, A. Hsu, "Smart image design for large image databases", *JVLC*, 3, 323-342, 1992.

[8] C. Colombo, A. Del Bimbo, P. Pala, "Semantics in Visual Information Retrieval", *IEEE Multimedia*, 6, 38-48, 1999.

[9] J.M. Corridoni, A. Del Bimbo, M. Mugnaini, P. Pala, F. Turco, "Pyramidal Retrieval by Color Perceptive Regions", *VL'97*, 209-215, 1997.

[10] C. Gurr, "Effective Diagrammatic Communication: ", *JVLC*, 10, 317-342, 1999.

[11] V. Haarslev, M. Wessel, "Querying GIS with Animated Spatial Sketches", *VL'97*, 201-208, 1997.

[12] R. Inder, N. Bianchi-Berthouze, T. Kato, "BK-DIME: A Software Framework for Kansei Filtering of Internet Material",. *IEEE SMC 1999*, vol.4, 358-363, 1999.

[13] A.K. Jain, A. Vallaya, "Image retrieval using color and shape", *Patt. Rec.*, 23(9), 1233-1244, 1996,

[14] T. Kurita,T. Kato, “Learning of Personal Visual Impression for Image Database Systems”, *ICDAR’93*, 547-552, 1993.

[15] L. Steels, F. Kaplan, "Spontaneous Lexicon Change", *COLING - ACL 98*, 1243-1249, 1998.

[16] D. Wang, H. Zeevat, " A syntax directed approach to picture semantics", in *Theory of Visual Languages*, K. Marriott and B. Meyer eds., Springer, 1998, 307-323

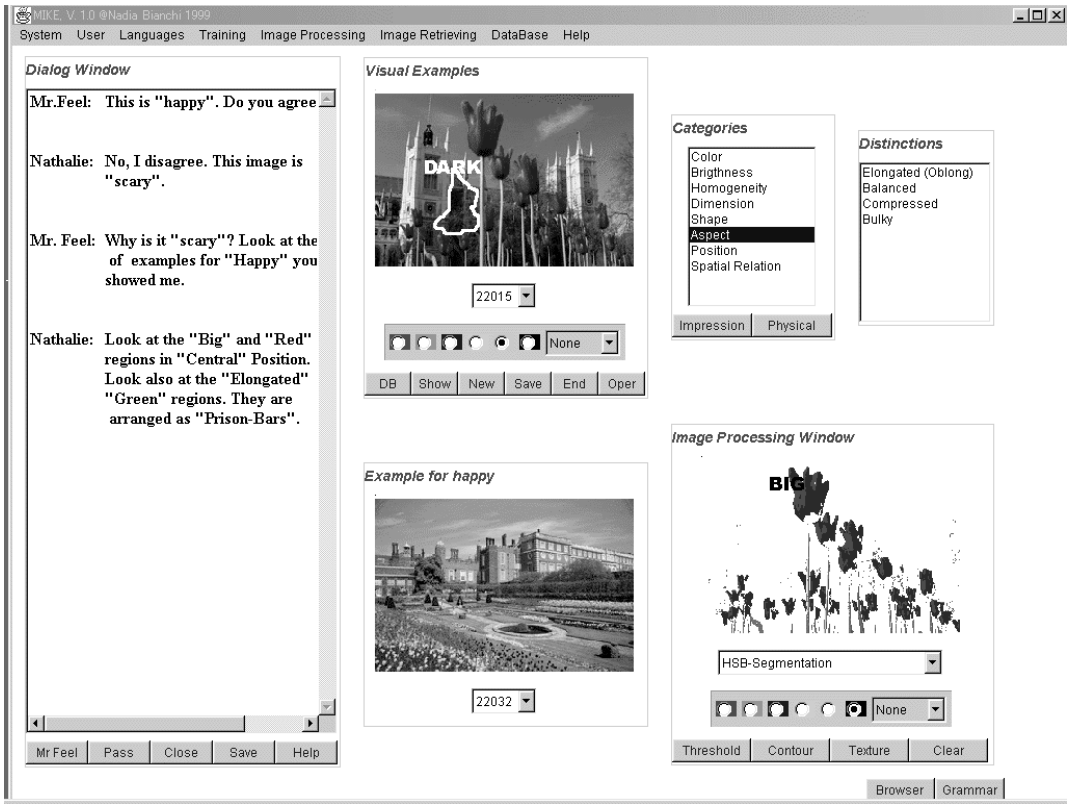

Figure 3. User and system discuss on the impression evoked by an image.