ADAPTIVE USER-DEFINED SIMILARITY MEASURE

J. Philippeau, P. Joly, J. Pinquier

IRIT UMR 5505 CNRS-INP-UT1-UPS
118 Route de Narbonne
31062 Toulouse Cedex 9, FRANCE

ABSTRACT

With the aim of audiovisual database consulting, the prospect of an interactive visual organization tool should be enviable. We consider a document as a visual entity in a computer screen. Thanks to an appropriate simplified GUI, a user can organize a few documents. Its notion of audiovisual similarity is relative to the Euclidean distance between entities: the more they are close, the more they are similar. The system infers a measure of similarity, that relies on the analysis of low-level features, and reorganizes the remaining of the database. This measure, based on support vector regression, conciliates human perception of audiovisual similarity and low-level automatically extracted data.

1. INTRODUCTION

We want to develop a system with respect to several principles:

1. Context-free user behavior: We do not want to focus on a particular trade (documentalist, montage specialist...). This precise point encouraged us to design an uncluttered dynamic interface, without any trade-oriented information, that is used for both similarity learning and result presentation.

2. Documents heterogeneity: Any kind of mono-media or multi-media document should be accessible in the same way (textual modality has not been explored yet).

3. Lack of descriptive values information: The field of semi-supervised audiovisual analysis we are interested in relies on low-level features extraction and modeling. We consider that descriptors we have are independent (so descriptive values computed on documents are), and that we do not know their behavior: Are they linear, logarithmic, etc...? What are their minimum and maximum values? Are they continuous or not?

4. Resulting application as generic as possible: Similarity is used in classification, identification and characterization tasks [2]. Our application has to cover the possible accomplishment of any task relying on those three aspects of the expression of similarity. We will use the global term of organization to evoke them. We want to create an interactive system that allows to learn a user-defined similarity and to organize a database with the same Graphical User Interface (GUI).

This problematic leads to focus on two points of view: a subjective one relying on user-defined audiovisual similarities, and an objective one that depends on low-level descriptive values.

2. GUI

2.1. Functionalities

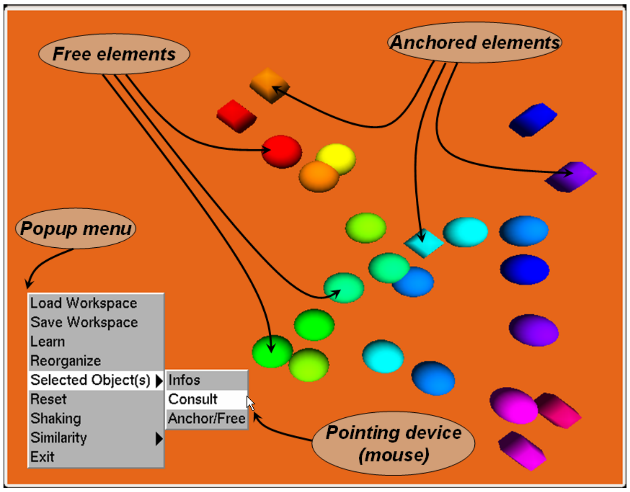

Fig. 1. Annotated screen shot of the Graphic User Interface.

- All functionalities usually implemented in a WIMP-like (Window, Icon, Menu, Pointing device) interface are present: single or multiple selection of elements, "drag and drop" process, etc...

We added some other functionalities: zoom and global translations movements of the graphical environment; content consulting for video or audio medias; possibility to change entities' representations (texture, color) in interests of readability; possibility to save and to load entities positions and computed similarity measure.

- The user can "anchor" the entities (drawn as diamonds in figure 1), opposed to free elements (drawn as spheres): anchored items are not able to move in space anymore.

- The user select several documents and click on the "Learn" button. Those document (composing the training set) are anchored to symbolize users trust in their placement. The learning process generates a similarity measure based on a set of descriptive values.

- By clicking on the Reorganize button, the system directly uses the learnt similarity measure on a set of selected entities. During the reorganization phase chosen entities are moving in the space until they reach a stability state.

2.2. Dynamic visual engine

We decided to consider visual entities as physical particles and to implement a "mass spring" dynamic physical model. We implemented this model with the four order Runge-Kutta algorithm, a temporal explicit integration schema known to be very accurate and well-behaved for a wide range of problems. The global set of documents is considered as a complete graph whose nodes are weighted particles and edges are springs.

Given an arbitrary time step, this algorithm computes the new position of a point with an approximation of its velocity, regarding all the implemented physical constraints which are the weight of a particle, an attraction force that is proportionate to the expected spring length / actual spring length ration, a moderation force adjoining the particle (proportionate to its velocity) that constraints the system to reach a stability state even if it does not exist and the strains of the springs.

3. LEARNING ENGINE

The main idea is to generate a behavioral model of a set of low-level descriptors. This model has to be representative enough of the arrangement made by the user in the visual interface.

We choose an early fusion strategy because of the heterogeneity of our features and the lake of information we have about them. We apply it inside a modality (audio or video) and between them.

We use the following Min-Max normalization method: each feature is scaled in a [0-1] range before being concatenated.

The Mean Square Error (or MSE) has been chosen as an indicator to process our similarity measure.

To summarize, each pair of documents, placed by the user in the visual space, generates a normalized distance value and a vector composed by normalized concatenated descriptive values. We choose a regressive model, based on ε-Support Vector Regression [3], to bind them.

Furthermore, an iterative concatenation process specifies which descriptors to keep, to constitute a good model. Here is a short explanation of the general idea :

* After the dynamic users organization phase, regression models are created (one model for each feature vector) by iteratively concatenating descriptive values, like a Sequential Forward Selection algorithm [1].

* We use the MSE to evaluate whether a regression is better than another.

* Iteration is done by concatenating the descriptive values that give the best results (regarding the MSE) at each algorithm step. This is done until the algorithm leads to a Loss of Performance (i.e. the best MSE performed at step p is lower than at the one at step (p+1)) or to an Information Redundancy (i.e. the best descriptor found at the actual step has already been chosen during the iterations).

* The reorganization phase consists in the application of the previously computed regression on a new subset of documents to generate a similarity matrix.

* Finally, the dynamic visualization engine generates at each time step a new distance matrix from the similarity matrix, until the global schema reaches a stability state.

4. CONCLUSION

Adjustment of appropriate evaluation tasks is currently being developed. Furthermore, extension of those tests on bigger corpora is in progress, and we plain to integrate textual modality as soon as possible.

Moreover, a very interesting aspect we are working on is the possibility to use our system to hierarchically navigate inside a database (from collections of TV broadcast to TV news, then to reports and lastly to video shots for example): because each level has its own set of specific descriptive values, it should be possible to automatically reorganize lower hierarchical levels while projecting the user-defined similarity expressed in an upper one.

5. REFERENCES

[1] P.A. Devijver and J. Kitter, "Pattern recognition: A statistical approach", in journal *Prentice Hall*, 1982.

[2] G. Bisson. "Why and How to Define a Similarity Measure for Object Based Representation Systems", in *Proceedings of 2nd international conference on building and sharing very large-scale knowledge bases (KBKS)*, IOS press, 236-246,

[3] C. Cortes and V. N. Vapnik, "Support vector networks", in journal *Machine Learning*, vol. 20, pp. 1–25, 1995.