Interactive Face Retrieval using Self-Organizing Maps

Pablo Navarrete and Javier Ruiz-del-Solar
Department of Electrical Engineering, Universidad de Chile.
Av. Tupper 2007, Santiago - CHILE
Email: {pnavarre, jruizd}@cec.uchile.cl

Abstract – An interactive face retrieval system that uses self-organizing maps and user feedback is described. The system solves some problems of related content-based image retrieval systems: non-existence of trivial high-level human descriptions of the images and the gap between the high-level descriptions and the low-level features used to index the images.

I. INTRODUCTION

Nowadays the massive access for information and the growing availability of digital image databases demand the existence of retrieval systems that can understand human high-level requests. For this reason content-based image retrieval is today an expanding discipline. The aim of this article is to tackle this problem in the specific case of face images. The content-based retrieval of faces has multiple applications that exploit existing face databases. In this sense, one of the most important tasks has always been the problem of searching a face without having an explicit image of it, but only its remembrance.

There are many alternatives in order to develop an efficient system for content-based face retrieval. In this work we have used the so-called *relevance feedback* approach. Under this approach previous human-computer interactions are employed to refine subsequent queries, which iteratively approximate the wishes of the user [1]. This idea is implemented using self-organizing maps. In particular, our system uses a tree-structured self-organizing map (TS-SOM) [2], for auto-organizing the face images in the database. Similar face images are located in neighbor positions of the TS-SOM. In order to be auto-organized, face images must be represented by feature vectors in the map. Based on one of the most successful approaches used in face recognition we decided to use PCA-projections for this task. Principal Component Analysis (PCA) projects face images onto a dimensional reduced space where the faces are well represented in a holistic sense.

In order to know in which part of the map the requested face is located, the system asks the user to select face images, which she considers are similar to the requested one, from a given set of face images. Then, the system shows the user new face images, which have neighbor positions in the map, respect to the ones selected for the user. The user and the retrieval system iterate until the interaction process converges, i.e. the requested face image is founded. A block diagram of our face retrieval system and its interaction with the user is shown in figure 1. In figure 4 is shown a real example of the interactive face-retrieval process.

As can be noted, our approach for the interactive content-based face retrieval is based on the following assumptions:

1) the face space forms a cluster in the whole image space,
2) PCA-projections give a suitable representation of the image space cluster formed by the face images, and
3) the PCA-representation together with the similarity measure used in the off-line TS-SOM training (usually Euclidean metric) is consistent with the criteria utilized by humans for the selection of similar faces.

The article is structured as follows. The face retrieval system implementation is described in section 2. In section 3 the properties of this system are studied based on simulation results. Finally, in section 4 some conclusions of this work are given.

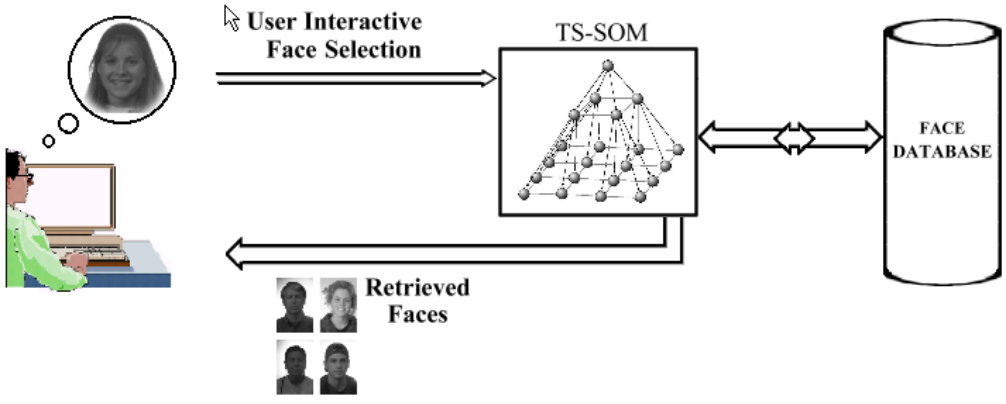

FIGURE 1. Block diagram of the face retrieval system.

II. FACE RETRIEVAL SYSTEM

A. Tree Structured SOM – TS-SOM

The TS-SOM is a tree structured vector quantization algorithm that uses self-organizing maps at each of its levels. In our system, as well as in PicSOM [1], all TS-SOM maps are two-dimensional. Figure 2 shows an example of a TS-SOM structure.

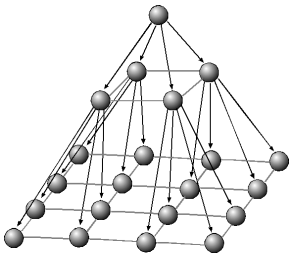

FIGURE 2. Example of a two-dimensional TS-SOM structure of 3 levels.

The TS-SOM training procedure begins with the highest map and it continues with the lowest ones. As in any Self-Organizing Map, each node (or neuron) has a weight-vector that represents a certain area of the high-dimensional input space. In addition, the TS-SOM structures allow connections between each node (the parent) and four nodes in the next level (the children), as shown in Figure 2. In order to determine the weight-vectors of all nodes at all TS-SOM levels (offline training), using the training samples (face images), the TS-SOM algorithm carries out the following iteration procedure:

0) In the first level (that has one node) the weight-vector is the mean of all the training samples.
1) In the other levels, weight-vectors of the previous level are copied into the children of the current level (initialization).
2) The centroid associated to each node is determined as the mean of the closest training samples. The set of closest samples is found based on limited search, that is searching only in the children nodes of the parent (best matching unit) of the previous level and the children of the parent's neighbors.
3) The weight-vectors are calculated as the mean of the neighbor centroids, weighted by the number of closest samples (of each node) and a kernel function that gives more importance to the closest neighbors (usually a Gaussian function).
4) If the weight-vectors do not change more than a given value, then the procedure continues with the next level in step 1). If not, then it goes into step 2).

In order to use a self-organizing map for a content-based image retrieval we must worry about the complexity involved in the training process. A standard SOM map needs a long time to train huge maps using large databases. TS-SOM was originally intended as a fast implementation of the SOM. In fact, the TS-SOM algorithm works much faster because of the limited updating in each training level. Also the search space on the underlying SOM levels is restricted to a predefined portion of the map, just below the best-matching unit on the above SOM. Then, the complexity of searches in a TS-SOM structure with N units is $O(\log_p N)$, with p the number of children per node. This property explains the reasons for using this structure in our face retrieval system.

B. Principal Components Analysis - PCA

PCA is a general method to identify the principal differences between signals and after that to make a dimensional reduction of them. In order to obtain the eigenfaces (face vectors in the reduced space), we need to obtain the projection axes in which exists the largest variance of the projected face images. Then, we repeat this procedure in the orthogonal space that is still uncovered, until we realize that there is no more variance to take into account. The theoretical solution of this problem is well known and is obtained by solving the eigensystem of the correlation matrix $\mathbf{R} \in R^{N \times N}$:

$$\mathbf{R} = \mathrm{E}\left\{(\mathbf{x} - \overline{\mathbf{x}})(\mathbf{x} - \overline{\mathbf{x}})^{\mathbf{t}}\right\} \qquad (1)$$

where $\mathbf{x}$ represent the normalized image vectors, $\overline{\mathbf{x}}$ is the mean face image, and N is the original vector image dimension. The eigenvectors of this system represent the projection axes or eigenfaces, and the eigenvalues represent the projection variance of the correspondent eigenface. Then by sorting the eigenfaces in descendent order of eigenvalues we have the successive projection axes that solve our problem.

The main problem is that $\mathbf{R} \in R^{N \times N}$ is too big for a reasonable practical implementation. We have a database of NT face images (the training set), and then we need to estimate the correlation matrix just by taking the corresponding averages in the training set. Let $\mathbf{X} = \left[(\mathbf{x}^1 - \overline{\mathbf{x}})(\mathbf{x}^2 - \overline{\mathbf{x}}) \cdots (\mathbf{x}^{NT} - \overline{\mathbf{x}})\right]$ be the matrix of the normalized training vectors. Then, the $\mathbf{R}$ estimator will be given by $\mathbf{R} = \mathbf{X}\mathbf{X}^{\mathrm{T}}$. We could say that the number of eigenfaces must be less than or equal to NT, because with NT training images all the variance must be projected into the hyperplane subtended by the training images. In other words the rank of $\mathbf{R}$ is less than or equal to NT. Thereafter they could have more null or negligible eigenvalues depending on the linear dependence of the vectors in the training set. In addition, the eigensystem of $\mathbf{X}^{\mathrm{T}}\mathbf{X} \in R^{NT \times NT}$ has the same non-zero eigenvalues of $\mathbf{R}$, because $\mathbf{X}\mathbf{X}^{\mathrm{T}}\mathbf{X}\,\mathbf{v}^k = \lambda_k\,\mathbf{X}\,\mathbf{v}^k$ represent both systems at the same time.

Now we can solve the reduced eigensystem of $\mathbf{X}^{\mathrm{T}}\mathbf{X} \in R^{NT \times NT}$. The correspondent eigenvalues are just the eigenvalues of the original system, and the eigenfaces are represented by $\mathbf{w}^k = \mathbf{X}\,\mathbf{v}^k$, and to be normalized they must be divided by $\sqrt{\lambda_k}$.

1. A random set of faces is presented to the user.

2. User interactive selection of faces.

3. System content-based face retrieval.

4. User analysis of retrieved faces.

- **4.a.** Requested face was found → Exit
- **4.b.** Similar faces were found → Go to 2
- **4.c.** No similar faces were found
  - **4.c.1.** User tired → Exit
  - **4.c.2.** User no tired (re initialization) → Go to 1

FIGURE 3. Interaction procedure between the retrieval system and users.

C. Content-based Retrieval System

In our work we have developed a system based in the PicSOM implementation [1]. In order to know in which part of the map the requested face is located, the system asks the user to select face images that she considers are similar to the requested one, from a given set of face images. Then, the system shows the user new face images, which have neighbor positions in the map, respect to the ones selected for the user. The user and the retrieval systems iterate until the interaction process converges. The interaction procedure between the retrieval system and the user is outlined in figure 3.

At first the system shows a random set of face images from the database. The user selects the retrieval face images most similar to the face she is looking for. With this information the system updates a so-called selection matrix, by increasing/decreasing those elements corresponding to the nodes on the TS-SOM net in which the face images selected/ignored are respectively located. A low-pass filter is then applied on the selection matrix in order to spread the positive/negative response. The new set of retrieval face images is selected by choosing images whose corresponding nodes have the highest values in the selection matrix. The hypothesis assumed under this approach is that the PCA-representation together with the similarity measure used in the off-line TS-SOM training (usually Euclidean metric) is consistent with the user-criteria used for the selection of similar faces. If this hypothesis is true, then the highest values in the selection matrix should move the searching process towards the zone in the TS-SOM map in which the searched face is located.

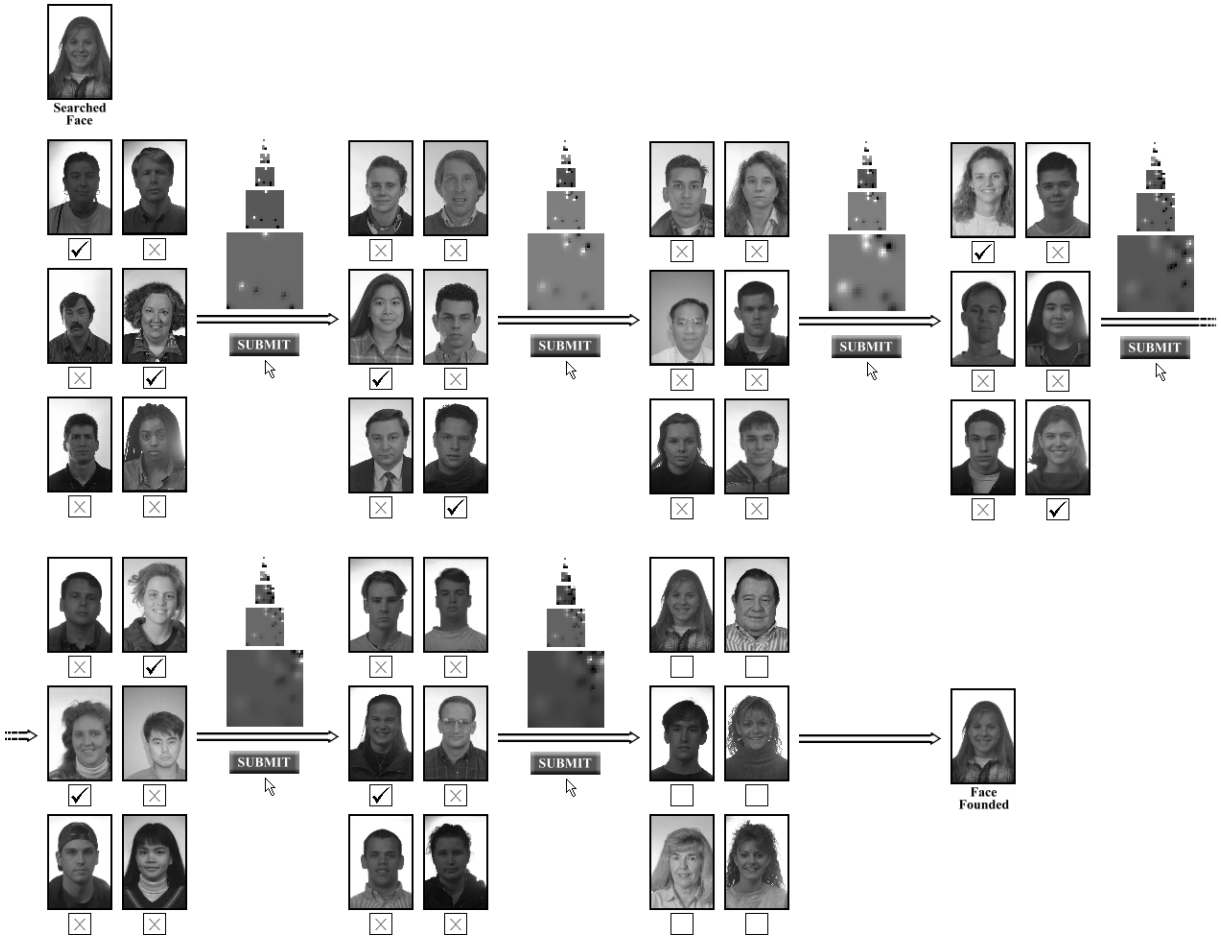

FIGURE 4. Simulation of the searching process using 684 face images of the FERET database. The system shows 6 face images per iteration. In this example the system shows the requested face image in the 6th iteration. The 6 maps between iterations, show the selection matrices for each level that determine the face images to be shown in the next iteration.

III. SIMULATIONS

In practical implementations the face retrieval system needs a large face image database in which several similar face images are available. In other case the user will have no reference faces in order to find the person she is searching for.

For testing our system we choose to use 684 face images of the FERET database [3]. We employed a TS-SOM with 6 levels and 32x32 nodes in its lowest level. In Figure 4 we show an example of a real searching process. In this example 6 face images are shown in each iteration to the user. After the selection process the selection matrices are computed and new images are chosen for the next iteration. These images are chosen by searching for face images not shown before, and closer to those points in which the selection matrices reach the highest values in all levels. Selected face images determine a set of zero-mean values that are summed to those points on selection matrices in which the selected face images are located. Then a low-pass filter is applied on each selection matrix in order to spread the selection values over each level.

IV. CONCLUSIONS

In this work a content-based face retrieval system that uses self-organizing maps and user feedback was presented. To build this system we assumed that face images could be founded in large databases by searching for similar faces in Self-Organizing Maps. Also we assumed that the PCA projection methods work as a suitable representation of the image space cluster formed by face images, as well as a representation consistent with human criteria of similar faces.

In our simulations we have realized that this kind of systems do work if large databases with several similar faces are used. As a future work we want to perform a detailed study of the characteristics of the proposed system.

ACKNOWLEDGEMENTS

This research was supported by the DID (U. de Chile) under Project ENL-2001/11 and by the join "Program of Scientific Cooperation" of CONICYT (Chile) and BMBF (Germany).

Portions of the research in this paper use the *FERET* database of facial images collected under the *FERET* program.

REFERENCES

[1] J. Laaksonen, M. Koskela, S. Laakso and E. Oja, “PicSOM – content-based image retrieval with self-organizing maps”, *Pattern Recognition Letters*, vol. 21, 1199-1207, 2000.

[2] P. Koikkalainen, Oja E., “Self-organizing hierarchical feature maps”, *Proc. Of 1990 Int. Joint Conf. on Neural Networks*, vol. II. IEEE, INNS, San Diego, CA, 1990.

[3] P. J. Phillips, H. Wechsler, J. Huang, and P. Rauss, “The FERET database and evaluation procedure for face recognition algorithms,” *Image and Vision Computing J.*, Vol. 16, no. 5, 295-306, 1998.