Incremental Face Recognition By Tagged Neural Cliques

Ehsan Sedgh Gooya

Institut Mines-Telecom
Telecom Bretagne
UMR CNRS 6285 Lab-STICC
Department of Electronic
Technopôle Brest Iroise-CS 83818
29238 Brest Cedex, France
Email: ehsan.sedghgooya@imt-atlantique.fr

Dominique Pastor

Institut Mines-Telecom
Telecom Bretagne
UMR CNRS 6285 Lab-STICC
Department of Signal and communication
Technopôle Brest Iroise-CS 83818
29238 Brest Cedex, France
Email: dominique.pastor@imt-atlantique.fr

***Abstract*—We present a system aimed at performing an incremental learning based on a neural network of tagged cliques for face recognition. A crucial component of the system is the network of neural tagged cliques. In its original version, cliques are a set of binary connections linking a set of fired neurons. Tagged cliques make it then possible to identify these cliques. The incremental learning is achieved through two phases: the first one is supervised by an oracle and the second one is automatic. Experimental results on the ORL (Olivetti Research Laboratory) face database pinpoint that incremental learning significantly reduces the number of features to store and yields substantial recognition rate improvement, in comparison with no incremental learning.**

***Keywords*–Face recognition; incremental learning; neural tagged cliques; SIFT (Scale-Invariant Feature Transform) features.**

I. Introduction

Developing brain-like systems has become a cutting-edge research topic in bio-inspired computational methodologies and approaches to address complex real-world problems to which traditional approaches are ineffective or infeasible.

Facing a new situation, human beings use their past experience to remember similar situations and enrich their knowledge. The purpose of this paper is to introduce a neural network system aimed at mimicking this behavior.

Basically, our approach is inspired by advances in cognitive science, such as [1], which led to the theory of dynamic memory, according to which the cognitive processes of understanding, memorization and training are based on the same memory structure. This structure is described by the *Organization Packets* and represented via knowledge representation schemata such as conceptual graphs and scripts. This structure is adapted to cope with new situations because "In the human memory, patterns are both a way of representing the knowledge organization and a way to express how this knowledge is used to understand, remember and make inferences." [2].

On the basis of the foregoing references in cognitive science, we hereafter propose a new incremental learning system based on neural network of tagged cliques for face recognition.

Regarding face recognition, much attention has been given to feature-based methods, such as SIFT (Scale-Invariant Feature Transform) [3], due to the fact that these descriptors remain invariant under rotation, scaling and variation in lightning condition. In the conventional face recognition method using local SIFT features [4] [5], SIFT features are extracted from all the faces of the database. Then, given a query face image, each feature extracted from that face is compared to those of each face contained in the database. A query feature is considered to match one of the database according to a certain threshold-based criterion. The face in the database with the largest number of matched descriptors is considered as the nearest face.

This new architecture relies on the neural network of tagged-cliques presented in [6], as a continuation of [7]–[11]. Cliques exhibit properties that are particularly relevant for incremental learning.

In Section II, we describe the proposed clique-based incremental learning system. In Section III, an implementation of the proposed system is described. Then the system is tested on ORL face database. Finally, we conclude in section IV.

II. Clique-based incremental learning system

A. Overview

The system we propose relies on two spaces: the knowledge space and the space of tagged cliques. These two spaces form what is hereafter called the knowledge structure. The knowledge space and the space of tagged-cliques are updated during the training via several processes. These processes, associated with the knowledge structure, are organized according to three phases during incremental learning (see Figures 1 and 2). These three phases are the initialization phase, the off-line phase and the on-line phase. Thanks to the initialization phase, the knowledge space and the space of tagged-cliques are created. The processes evoked above are then used to update incrementally the knowledge space and the space of tagged-cliques. The processes involved in the off-line phase are: recall, verification, adaptation, evolution and memorization. This updating of the knowledge structure is firstly performed off-line, during which an oracle supervises the learning. During the on-line phase, the updating is carried out automatically without any supervision. In this phase, the verification process is replaced by a validation process.

Let us now specify the approach with respect to face recognition. Basically, face recognition is aimed at determining a person identity, given a face image of this person. This objective requires prior knowledge of the person identity, on the basis of one or several images of this person's face. The system acquires this prior knowledge by a training phase based on a training database of images. A test base of images is then usually employed to assess the performance of the face recognition system.

In the incremental-training system proposed in this paper, one image is randomly chosen in the training database so as to initialize the knowledge structure and store the identity of the several persons to recognize. Afterwards, during the off-line phase, the system is upgraded under the supervision of the oracle. During the on-line phase, the system estimates itself the identity of the input image before updating the knowledge structure.

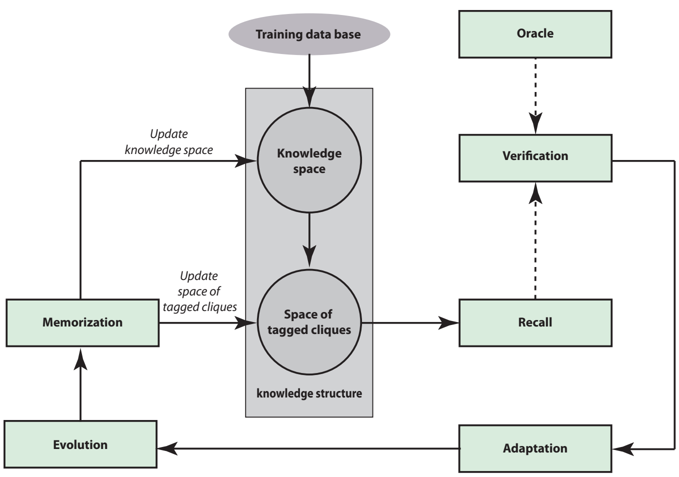

Figure 1. Off-line training

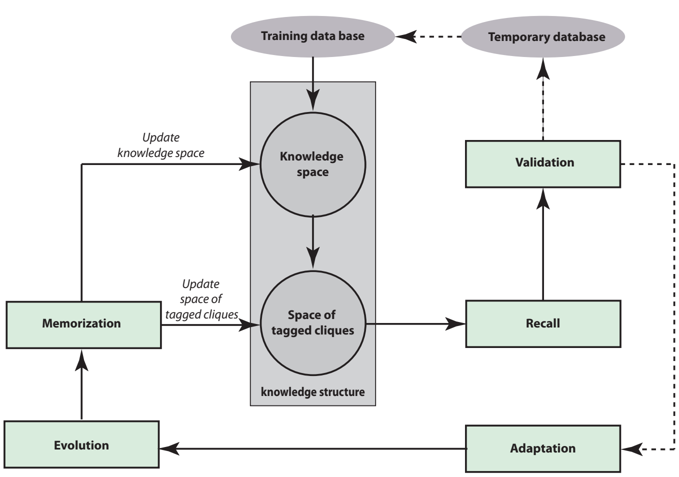

Figure 2. On-line training

III. Implementation

A. System initialization

Let us denote by $\{I_i\}_{i=1}^{L}$ the set of L images that are available in the training database. In the initialization phase, we begin by randomly selecting one single image for each person k represented in the training database. By so proceeding, we obtain c images $\{J_k\}_{k=1}^{c}$, where c is the number of persons to cope with. This set of images is used to create and initialize the knowledge structure as follows.

The space of tagged-cliques: This space is created as proposed in [6]. More specifically, we consider n neurons. This set of neurons is split into two non-intersecting clusters. Cluster #1 contains d neurons and cluster #2 involves c neurons so that: $n = d + c$. The d neurons of cluster #1 are indexed from 1 to d and the c neurons of cluster #2 are indexed from $d + 1$ to n. The space of tagged-cliques is then constructed to store the c gallery-vectors $\mathbf{g}_k$ for $k = 1, 2, \ldots, c$ corresponding to the c persons to recognize. This construction is carried out according to the several steps described below.

Initialization of the knowledge space: We calculate the set F_k of the local features of each given image J_k [3], [4], [12]. Let us suppose that $F_k = \{F_k^1, \ldots, F_k^m\}$, where the F_k^j's are the local SIFT features and m is the number of these local SIFT features extracted from the image. Note that m may differ from one image to another. To each F_k^j, we associate a neuron n_k^j in cluster #1. This choice may be random, but constrained so as to be one-to-one for person k. Let Ψ stand for this correspondence so that $\Psi(F_k^j) = n_k^j$. We then create the one-to-one correspondence that assigns to each $F_k = \{F_k^1, \ldots, F_k^m\}$ the set $N_k = \{n_k^1, \ldots, n_k^m\}$. This correspondence can be represented by the set of pairs: $\mathbf{D}_k = (F_k; N_k)$. This set $\mathfrak{S} = \{\mathbf{D}_k : k = 1, 2, \ldots, c\}$ is the initial knowledge space and our purpose is then to upgrade this knowledge space.

Initialization of the space of tagged-cliques: After determining and storing $\mathbf{D}_k$, the vector N_k of neuron indexes is stored in the space of tagged-cliques by proceeding as follows.

1) We define the binary pattern $\boldsymbol{x}_k$ with dimension d ($\boldsymbol{x}_k \in \{0;1\}^d$) by setting:
$$(\boldsymbol{x}_k)_i = \begin{cases} 1 & \text{if } i \in N_k \\ 0 & \text{otherwise} \end{cases} \tag{1}$$
2) We associate to $\boldsymbol{x}_k$ the k^{th} element e_k of the canonical basis of $\mathbb{R}^c$, which can be regarded as a very basic full disjunctive coding:
$$(\boldsymbol{e}_k)_i = \begin{cases} 1 & \text{if } i = k \\ 0 & \text{otherwise} \end{cases}$$
In other words, vector e_k represents the identity of person k.
3) The gallery-vector $\boldsymbol{g}_k$ is then defined by
$$\boldsymbol{g}_k = \begin{pmatrix} \boldsymbol{x}_k \\ \boldsymbol{e}_k \end{pmatrix} \tag{2}$$

4) The storage of the gallery-vectors $\boldsymbol{g}_k$ is performed by calculating the adjacency matrix:

$$\boldsymbol{W} = \bigvee_k \boldsymbol{g}_k \, \boldsymbol{g}_k^T \tag{3}$$

where $(\cdot)^T$ is the standard transpose operator.

B. Off-line training

After initializing the knowledge space and the space of tagged-cliques as described above for only one single image per person to recognize, the off-line phase is engaged. This off-line updating makes it possible to reduce the number of features to store in the knowledge base and the number of possibly contentious connections in the space of tagged-cliques. Only features of interest will be added in the knowledge structure during the updating.

The off-line phase is applied to the images

$$\bar{I} \; = \{I_i\}_{i=1}^{L} \setminus \{J_k\}_{k=1}^{c}$$

that were not selected in the initialization phase. The several processes (recall, verification, adaptation, evolution and memorization) are performed as follows.

1) Recall process: Given $\bar{I}_j \in \bar{I}$, we calculate as above the feature vector $\bar{F}_j$ by extracting the local SIFT vector features in $\bar{I}_j$. We then look for the closest feature vector $\bar{F}$ in the knowledge space via a simple L_2 minimization. This vector feature $\bar{F}$ is associated with a set of neurons $\bar{N}$. We derive the input pattern $\boldsymbol{x}$ associated with $\bar{F}$ and $\bar{N}$ via (1) with $N_k = \bar{N}$. According to [12], we then use Algorithm 1 with

$$f(\mathbf{v})_i = \begin{cases} 1 & \text{if } \mathbf{v}_i = \max_j \mathbf{v}_j \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

to estimate the pattern and identity corresponding to $\bar{F}$ and its associated set $\bar{N}$ of neurons. In this algorithm, $\pi_{\mathbf{x}}(\mathbf{v})$ (resp. $\pi_{\mathbf{e}}(\mathbf{v})$) extracts the vector made of the first (resp. last) d (resp. c) coordinates of $\mathbf{v} \in \mathbb{R}^n$.

Algorithm 1: Recall algorithm by neural network of tagged cliques

---

Input: Input pattern $\boldsymbol{x}$ and adjacency matrix $\boldsymbol{W}$
Output: $\widehat{\boldsymbol{e}}_k$, the class indicator vector estimated for $\mathbf{x}$

1 $\widehat{\boldsymbol{x}} = \pi_{\boldsymbol{x}} \left(f(\boldsymbol{W} \begin{pmatrix} \boldsymbol{x} \\ \mathbf{0}^{\mathrm{c}} \end{pmatrix}) \right)$;

2 $\widehat{\boldsymbol{e}}_k = \pi_{\boldsymbol{e}} \left(f(\boldsymbol{W} \begin{pmatrix} \widehat{\boldsymbol{x}} \\ \mathbf{0}^{\mathrm{c}} \end{pmatrix}) \right)$;

---

2) Verification process: During the off-line training phase, we suppose that the identity of the face image is known. We thus propose a verification process by oracle. By thus proceeding, the identity returned by the system is verified and compared to the supervisor knowledge during the off-line training phase. This verification avoids that the system makes erroneous identifications, which is probable as long as the system has not acquired enough knowledge to allow for automatic identification.

3) Adaptation process: At the end of the recall and verification processes, the identity of the face query image is determined. For a given identity k issued from these processes and validated by the supervisor, the purpose of the adaptation process is to discriminate the knowledge already stored in the knowledge structure for person k from that brought by the new image.

During the adaptation process, the neuron indexes used to encode person k as a clique are determined from the space of tagged cliques by algorithm 2, where $\mathbf{0}^d$ is the zero vector with dimension d.

Algorithm 2: Algorithm used to retrieve pattern $\boldsymbol{x}_k$ when $\boldsymbol{e}_k$ is known.

Input: $\boldsymbol{e}_k$, the class indicator vector
Output: $\boldsymbol{x}_k$, pattern associated to $\boldsymbol{e}_k$

1 $\boldsymbol{x}_k = \pi_x \left(W \begin{pmatrix} \mathbf{0}^d \\ \boldsymbol{e}^k \end{pmatrix} \right)$

At the end of the adaptation process, we obtain pattern $\boldsymbol{x}_k$. By inverting (1), we then determine the set N_k of neurons corresponding to $\boldsymbol{x}_k$. The set N_k can be regarded as the knowledge already stored for person k in the space of tagged-cliques. The new knowledge brought by a new image of person k is collected in a set denoted N^* and determined by:

$$N^* = \bar{N} \setminus N_k \tag{5}$$

The new features that can be associated with person k are then given by $F^* = \Psi^{-1}(N^*)$, where Ψ is defined in Section III-A. In order to maintain the one-to-one correspondence between features and neurons for person k, the neurons in N^* are replaced by new randomly selected neurons in cluster #1 to form a new set of neurons. This new set is still denoted N^* in what follows and can then be associated univoquely to F^* so as to form the new pair:

$$\mathbf{D}^* = (F^*; N^*) \,. \tag{6}$$

4) Evolution process: For person k, the adaptation process has separated new pieces of knowledge brought by a new image of k and knowledge already stored in the knowledge structure. The evolution process aims to combine these two types of information, namely the new pieces of knowledge and those already stored in the system. The combination then amounts to creating the new pattern $\boldsymbol{x}_k^* \in \{0;1\}^d$ as follows:

$$(\boldsymbol{x}_k^*)_i = \begin{cases} 1 & \text{if } i \in N_k \cup N^* \\ 0 & \text{otherwise} \end{cases}$$

In addition, the new gallery-vector $\boldsymbol{g}_k^*$ calculated according to

$$\boldsymbol{g}_k^* = \begin{pmatrix} \boldsymbol{x}_k^* \\ \boldsymbol{e}_k \end{pmatrix}$$

where $\boldsymbol{e}_k$ is the k^{th} element of the canonical basis in $\mathbb{R}^c$.

5) Memorization process: The memorization process has the role of storing the updated pattern $\boldsymbol{g}_k^*$ in the space of tagged cliques and adding the new identified features to the knowledge space. More precisely, updating the knowledge space is simply performed by adding $\mathbf{D}^*$ to the knowledge space by making:

$$\mathfrak{S} \leftarrow \mathfrak{S} \cup \{\mathbf{D}^*\}.$$

Regarding incremental learning in the space of tagged cliques, the new gallery-vector $\boldsymbol{g}_k^*$ is stored as a tagged clique as:

$$\boldsymbol{W} \leftarrow \boldsymbol{W} \bigvee \boldsymbol{g}_k^* \left(\boldsymbol{g}_k^*\right)^T$$

During the off-line phase, the update is performed for all the face images in the training database. This update will continue automatically and without the supervision of the oracle during the on-line phase.

C. On-line training

At the end of the off-line training phase, by engaging the testing database, the system continues to update automatically the knowledge structure without any supervision of the oracle. The processes involved in the on-line training phase are *Recall*, *Validation*, *Adaptation*, *Evolution* and *Memorization.*

During the on-line training phase, the recall, adaptation, evolution and memorization processes are the same as those used by the off-line training phase described in Section III-B.

Without supervision, the system may incorporate false information to persons. To avoid this, the system must reject query images whose identification is uncertain. Deciding whether a face query image must be rejected or not is performed by the *Validation* process, which replaces the verification process of the off-line training phase.

1) Validation process: At a given time t, accepting or not new information provided by a query face image I is the task of the validation process. This process determines the score Φ assigned to the recognition of I by:

$$\Phi = \pi_e \left(W \begin{pmatrix} \boldsymbol{x} \\ \mathbf{0}^c \end{pmatrix} \right), \tag{7}$$

where $\boldsymbol{x}$ is the input pattern calculated according to Eq. (1). The i^{th} coordinate of Φ represents the number of connections between the i^{th} neuron in cluster #2 and the neurons in cluster #1 that are associated with $\boldsymbol{x}$. The idea is to validate and thus incorporate the new information brought by I if only one single neuron in cluster #2 has received a significantly larger number of votes than any other neuron. The significance of the number of votes is determined by deriving the sorted values of Φ in descending order. More specifically:

- Let $\bar{\Phi} = (\Phi_{(1)}, \Phi_{(2)}, \cdots, \Phi_{(c)})$ be the sequence of the values of Φ sorted in descending order: $\Phi_{(1)} \geqslant \Phi_{(2)} \geqslant \ldots \geqslant \Phi_{(c)}$.
- Set up the test:

$$\Gamma = \begin{cases} 1 & \text{if } card(C) = 1 \\ 0 & \text{otherwise} \end{cases}$$

where $C = \{(1), (2), \cdots, \arg\max_{1 \leq i \leq c} \bar{\Phi}'\}$ and $\bar{\Phi}'$ is the derivative of $\bar{\Phi}$.

As a result, if the test decision Γ returns 1, the query face image is used to enrich the knowledge structure by the processes of the on-line training phase. At a given time t, if the test decision Γ returns 0, this image is sent to a temporary database. When the on-line training phase is completed, the images collected in the temporary database are re-injected into the system for a new re-evaluation.

D. Experimental results on the ORL database

We tested the incremental learning system described above on the Olivetti and Oracle Research Laboratory (ORL) database. There are 10 different images for each of the 40 distinct subjects. For some subjects, the images were taken at different times, varying lighting, with different facial expressions (open / closed eyes, smiling / not smiling), facial details (glasses / no glasses) and head pose (tilting and rotation up to 20 degrees). All the images were taken against a dark homogeneous background.

These experimentation make it possible to better assess the effect of the oracle's supervision. This supervision is basically parametrized by the number K of images used during the off-line training phase. More specifically, if only one image per person is used to initialize the training, $K-1$ images among the remaining ones in the database for a given person will be employed. After off-line upgrading of the knowledge structure, the on-line training is engaged on the images remaining in the database. Once the on-line training is terminated, we assess the ability of the system to recognize the identity of the persons whose images are present in the test database. For every given K, the face recognition performance measurements of the system are given in Table I by averaging the results over 10 different tests where the images during the training are randomly chosen for every test. By so proceeding, we follow standard recommendations of the literature on the topic.

TABLE I. FACE RECOGNITION RATES OBTAINED WITH AND WITHOUT INCREMENTAL TRAINING

Method	**Number of training images**			
	K = 5	K = 6	K = 7	K = 8
Static leaning	98.82	99.55	99.71	99.88
Incremental learning	**98.91**	**99.6**	**99.91**	**100**

According to these results, it turns out that from $K = 8$ onwards, the system commits no error to recognize the images of the test database. In any case, even when the face recognition rate is not 100%, incremental learning always yields performance improvement.

We can also consider the number of features stored in the knowledge space. We expect that incremental learning also optimizes this number. This is actually the case as shown by Figure 3. The number of features stored in the knowledge space is significantly lesser than that obtained without incremental learning.

IV. Conclusion

We have presented an approach that performs face recognition after incremental learning on the basis of

 ISBN: 978-1-61208-531-9

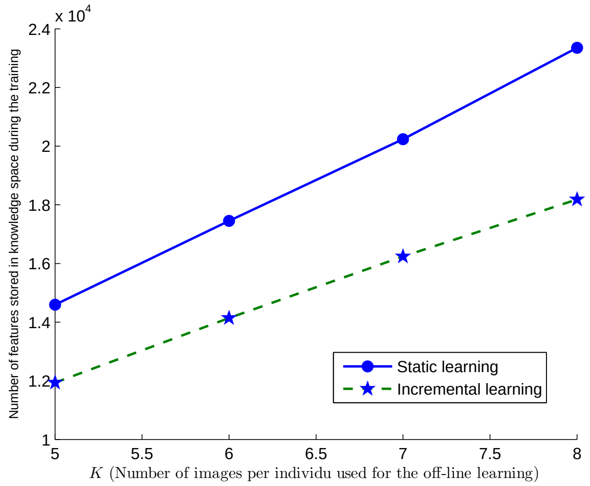

Figure 3. The number of features stored in knowledge space during the training

a neural network of tagged cliques. This system is an extension of the face recognition system introduced in [12]. Nevertheless, the reader will easily notice that the neural network of tagged cliques could certainly be replaced by other types of classification. For instance, features could be stored without any coding and exhaustive search would even be thinkable. However, networks of neural cliques present two fundamental advantages. First, the coding and decoding processes are fast and the storage of a new clique is performed independently of all the cliques previously stored.

The system proposed in this paper is then capable of updating its knowledge structure incrementally, first via a supervised phase and then automatically. Experimental results on the ORL database enhance the relevance of the incremental approach, which makes it possible to optimize the number of features stored and yield face recognition rates better than that obtained without incremental learning.

References

[1] C. Riesbeck and R. Schank, Inside case-based reasoning. Psychology Press, 2013.

[2] J. Richard, C. Bonnet, R. Ghiglione, M. Bromberg, J. Beavois, and W. Doise, “Traité de psychologie cognitive: cognition, représentation, communication,” 1990.

[3] D. Lowe, “Object recognition from local scale-invariant features,” in Computer vision, 1999. The proceedings of the seventh IEEE international conference on, vol. 2. Ieee, 1999, pp. 1150–1157.

[4] M. Aly, “Face recognition using sift features,” CNS/Bi/EE report, vol. 186, 2006.

[5] L. Lenc and P. Král, “Novel matching methods for automatic face recognition using sift,” Artificial Intelligence Applications and Innovations, 2012, pp. 254–263.

[6] S. Larroque, E. Sedgh Gooya, V. Gripon, and D. Pastor, “Using tags to improve diversity of sparse associative memories,” in Proceedings of Cognitive, March 2015, pp. 1–7.

[7] V. Gripon and C. Berrou, “Sparse neural networks with large learning diversity,” Neural Networks, IEEE Transactions on, vol. 22, no. 7, 2011, pp. 1087–1096.

[8] B. Kamary Aliabadi, C. Berrou, V. Gripon, and X. Jiang, “Learning sparse messages in networks of neural cliques,” arXiv preprint arXiv:1208.4009, 2012.

[9] X. Jiang, V. Gripon, and C. Berrou, “Learning long sequences in binary neural networks,” in COGNITIVE 2012, The Fourth International Conference on Advanced Cognitive Technologies and Applications, 2012, pp. 165–170.

[10] R. Danilo, H. N. Wouafo, C. Chavet, and P. Coussy, “Associative memory based on clustered neural networks: improved model and architecture for oriented edge detection,” in Conference on Design & Architectures for Signal & Image Processing, 2016.

[11] D. Ferro, V. Gripon, and X. Jiang, “Nearest neighbour search using binary neural networks,” in Proceedings of IJCNN, July 2016.

[12] E. Sedgh Gooya, D. Pastor, and V. Gripon, “Automatic face recognition using sift and networks of tagged neural cliques,” in Cognitive 2015, The Seventh International Conference on Advanced Cognitive Technologies and Applications, 2015.