

# Réduction bi-directionnelle d'images - Vers une méthode d'extraction de caractéristiques multi-niveaux

Marc Joliveau

CIRRELT- Université de Montréal  
C.P. 6128, succ. Centre-Ville, Montréal, Québec, Canada H3C 3J7  
marc.joliveau@cirrelt.ca

**Résumé.** Inspiré des performances du cerveau humain à identifier les éléments par la vue, le problème de la réduction de la dimension dans le domaine de la perception visuelle consiste à extraire une quantité réduite des caractéristiques d'un ensemble d'images afin de les identifier.

Ce papier présente une approche innovante bi-directionnelle d'extraction de caractéristiques d'images fondée sur l'utilisation partielle d'une méthode spatio-temporelle. Les expériences numériques appliquées sur 70000 images représentant des chiffres écrits à la main ainsi que sur 698 images illustrant un visage sous différentes postures démontrent l'efficacité de notre approche à fortement réduire la dimension tout en conservant les relations intelligibles entre les objets des données, permettant même d'obtenir une meilleure classification à partir des versions réduites des images qu'à partir des versions originales.

## 1 Introduction

Durant les dernières décennies, le problème de la réduction de la dimension, i.e., trouver une structure intelligible de faible dimension sous-jacente à des observations de dimension élevée, a capté beaucoup d'attention auprès de la communauté scientifique. Dans la même période, l'importance grandissante prise par les fichiers multimedia comme les images, les vidéos ou les fichiers de son a grandement contribué à l'extension des bases de données. Ces fichiers, et particulièrement les images, interviennent dans une grande variété de domaines tels que la médecine, le trafic, la sécurité, etc. ... Réduire les images d'une base de données est généralement affilié au problème de perception visuelle (*visual perception*) qui consiste à extraire une quantité réduite des caractéristiques perceptuelles d'un ensemble d'images afin de les identifier.

Ce papier introduit une approche innovante afin de réduire la dimension des images d'une base de données tout en conservant au mieux les relations entre elles. L'approche proposée se fonde sur l'adaptation partielle à des images d'une méthode de réduction de données spatio-temporelles : l'estimateur STPCA (pour *Space Time Principal Component Analysis*) introduit dans Joliveau (2008). Comme démontré dans les validations numériques, notre méthode d'extraction des caractéristiques se révèle particulièrement efficace avec les images, notamment grâce à son fonctionnement bi-directionnel qui consiste à appliquer une réduction des images simultanément sur leur largeur et leur longueur.

L'estimateur STPCA a été exploité dans différents contextes avec des objectifs variés. Tout d'abord, les auteurs de Joliveau et DeVuyst (2007) ont proposé une variante de l'estimateur STPCA prenant en compte la présence de valeurs manquantes avec des pertes de performance très limitées. Différentes applications de l'estimateur STPCA sur des données de trafic afin d'interagir avec des Systèmes Inteligents de Transport (*Intelligent Transportation Systems*) ont également été présentées dans Joliveau et DeVuyst (2008); Joliveau (2008); Bauzer-Medeiros et al. (2008), concernant notamment l'extraction de motifs spatio-temporels de situations typiques et atypiques afin d'émettre des prédictions sur les comportements atypiques du trafic à travers un réseau routier.

Ce papier, se focalise sur l'utilisation innovante du descripteur STPCA, un état intermédiaire lors du calcul de l'estimateur STPCA, afin de réduire la dimension d'images. Dans le but de valider notre approche, différentes expériences numériques sont appliquées sur les bases de données *Mnist* et *faces* contenant respectivement 70000 images de chiffres et 698 images illustrant un visage. La suite de ce papier est organisée de la manière suivante. La section 2 offre un bref aperçu des différentes méthodes de référence utilisées pour la réduction d'images, alors que la section 3 revient en détails sur l'intégralité du processus STPCA. Les sections 4 et 5 illustrent respectivement les validations numériques de notre approche ainsi que ses relations aux méthodes existantes. Finalement, la section 6 conclue le papier et indique quelques futures directions de recherches envisageables.

## 2 Méthodes d'extraction des caractéristiques

Lorsqu'on cherche à réduire la dimension d'une base de données constituée d'image, la méthodologie suivie est généralement d'estimer la géométrie intrinsèque de ces données. Pour cela, on utilise principalement des méthodes d'extraction des caractéristiques (*feature extractors*). Le principe des méthodes d'extraction des caractéristiques est de projeter les objets des données (dans notre cas les images) sur un hyperplan de dimension réduite tel que deux objets similaires sont représentés par des points proches, et deux objets dissemblables par des points éloignés.

Les techniques classiques d'extraction des caractéristiques concernent principalement l'Analyse en Composantes Principale (*Principal Component Analysis*) ainsi que l'Echelonnement Multidimensionnel (*Multidimensional Scaling*). L'Analyse en Composantes Principales, introduite par Jolliffe (1986), applique une décomposition en valeurs singulières aux données. Cette méthode est décrite plus en détails ultérieurement dans le papier car elle est à la base de l'approche par STPCA. L'Echelonnement Multidimensionnel a d'abord été introduit par Richardson (1938) dans le domaine de la psychologie, puis a été adapté par Kruskal et Wish (1978) afin d'être appliqué au secteur de l'informatique. Cette approche peut être vue comme l'adaptation d'une Analyse en Composantes Principales qui considère une mesure de similitude ou de préférence entre les objets des données (e.g., une mesure de distance) plutôt que les objets eux-même.

Plus récemment, de nouvelles méthodes telles qu'Isomap et l'algorithme LLE ont été introduites. Contrairement aux méthodes classiques, celles-ci sont capables d'estimer le degré de liberté non-linéaire sous-jacent à des observations naturelles complexes (e.g., des images de chiffres écrit à la main). Cette particularité fait de ces approches des méthodes non-linéaires

de réduction de la dimension. Afin d’extraire les caractéristiques de données, Isomap, introduit par Tenenbaum et al. (2000), applique un échelonnement multidimensionnel à partir de la matrice des distances géodésiques entre les objet des données, alors que l’algorithme LLE (pour *Locally Linear Embeddings*), proposé par Roweis et Saul (2000), utilise une approche locale par voisinage pondéré.

### 3 L’approche STPCA

Dans Joliveau (2008), les auteurs ont introduit l’estimateur STPCA (pour *Space-Time Principal Component Analysis*), un nouvel outil pour résumer des séries spatio-temporelles. Dans ce contexte, l’approche STPCA se fonde sur l’application simultanée d’une Analyse en Composantes Principales dans les deux directions spatiales et temporelles. Cette section présente une description détaillée de l’approche STPCA et relate de l’utilisation de cette méthode dans un contexte de réduction d’images.

#### 3.1 Conditions sur les données pour la réduction d’images

Jusqu’à aujourd’hui, l’approche STPCA a été introduite dans un contexte spatio-temporel, afin de traiter des séries temporelles de trafic routier issues de capteurs et non pour traiter des images. Dans ce contexte, l’application de cette méthode suppose que les données sont collectées lors de  $N$  réalisations (e.g., des jours) à partir de  $I$  différentes localisations (e.g., des capteurs géoréférencés) à des instants identiques pour toutes les localisations sur une durée totale  $J$ . Ainsi, les données collectées lors d’une réalisation  $n$  peuvent être stockées dans une matrice  $\mathbf{X}_n$  de taille  $I \times J$ . Chaque ligne de  $\mathbf{X}_n$  correspond à une localisation et chaque colonne à un instant de mesure.

Afin d’utiliser l’approche STPCA pour réduire des images en niveau de gris, il suffit de considérer chaque image des données comme une réalisation. En effet, une image  $n$  peut aisément être représentée sous la forme d’une matrice  $\mathbf{X}_n$  contenant le niveau de gris de chacun de ses pixels. En suivant cette analogie,  $N$  correspond donc maintenant au nombre d’images contenues dans les données,  $I$  au nombre de lignes de pixels des images, et  $J$  à son nombre de colonnes de pixels. Comme l’approche STPCA nécessite que la taille des matrices de réalisations  $\mathbf{X}_n$  soit identique pour toutes les réalisations  $n = 1 \dots N$ , la seule condition sur les données à respecter pour utiliser cette approche afin de réduire des images est que celles-ci doivent partager la même résolution.

Dans le cas d’images en couleurs, la procédure est légèrement différente. Chaque image  $n$  n’est plus représentée par une unique matrice  $\mathbf{X}_n$  mais par trois matrices de taille  $I \times J$  :

- $\mathbf{X}_n^R$  : qui contient les niveaux de rouge de chaque pixel de  $n$  ;
- $\mathbf{X}_n^G$  : qui contient les niveaux de vert de chaque pixel de  $n$  ;
- $\mathbf{X}_n^B$  : qui contient les niveaux de bleu de chaque pixel de  $n$ .

Afin de réduire des images en couleurs, l’approche STPCA doit être appliquées indépendamment sur chacun des trois sous-ensembles  $\mathcal{R}$ ,  $\mathcal{G}$  et  $\mathcal{B}$  tels que :

$$\begin{aligned}\mathcal{R} &= \{\mathbf{X}_n^R, n = 1 \dots N\}, \\ \mathcal{G} &= \{\mathbf{X}_n^G, n = 1 \dots N\}, \\ \mathcal{B} &= \{\mathbf{X}_n^B, n = 1 \dots N\}.\end{aligned}$$

### 3.2 Définition de l'approche STPCA pour la réduction d'images

L'approche STPCA, qui considère à la fois le descripteur STPCA et l'estimateur STPCA, se décompose en trois étapes :

1. Assemblage des matrices de réalisation horizontalement (pour une analyse sur la largeur des images) dans une unique matrice  $\mathbf{Y}$ , et verticalement (pour une analyse sur la longueur des images) dans une unique matrice  $\mathbf{Z}$ . Selon cette définition, la matrice  $\mathbf{Y}$  représente l'image obtenue par la concaténation de toutes les images des données les unes à côté des autres, alors que la matrice  $\mathbf{Z}$  représente l'image obtenue par la concaténation de toutes les images des données les unes au dessus des autres. Ainsi, la matrice  $\mathbf{Y}$  est composée de  $I$  lignes et de  $(J \times N)$  colonnes alors que la matrice  $\mathbf{Z}$  contient  $(I \times N)$  lignes et  $J$  colonnes.
2. Décomposition en valeurs singulières des matrices  $\mathbf{Y}$  et  $\mathbf{Z}$ . Pour la matrice de Gram  $\mathbf{G}^{lar} = \mathbf{Y}\mathbf{Y}^T$ , on calcule les  $K$  premiers vecteurs propres  $(\Psi_k)_{k=1\dots K}$ , avec  $K \ll I$ , puis on les mémorise dans une matrice  $\mathbf{P}$  :

$$\mathbf{P} = \text{col}(\Psi_1, \Psi_2, \dots, \Psi_K).$$

Pour la matrice de Gram  $\mathbf{G}^{lon} = \mathbf{Z}^T\mathbf{Z}$ , on calcule les  $L$  premiers vecteurs propres  $(\Phi_l)_{l=1\dots L}$ , avec  $L \ll J$ , puis on les mémorise dans une matrice  $\mathbf{Q}$  :

$$\mathbf{Q} = \text{col}(\Phi_1, \Phi_2, \dots, \Phi_L).$$

3. Détermination de la matrice d'ordre réduit des coefficients, ou descripteur STPCA, de chaque réalisation  $n$  donnée par :

$$\tilde{\mathbf{X}}_n = \mathbf{P}^T \mathbf{X}_n \mathbf{Q},$$

de taille  $K \times L$  où  $K$  et  $L$  sont choisis pour être petits. Cette étape intermédiaire permet finalement de calculer l'estimateur STPCA d'une réalisation  $n$  défini par :

$$\hat{\mathbf{X}}_n = \mathbf{P} \tilde{\mathbf{X}}_n \mathbf{Q}^T = \mathbf{P}\mathbf{P}^T \mathbf{X}_n \mathbf{Q}\mathbf{Q}^T.$$

Jusqu'à aujourd'hui, toutes les applications fondées sur l'approche STPCA se focalisent uniquement sur l'estimateur STPCA  $\hat{\mathbf{X}}_n$  afin de résumer efficacement des données, estimer des valeurs manquantes, ou extraire des connaissances (e.g., motifs spatio-temporels). Cependant, l'état intermédiaire calculé par le descripteur STPCA  $\tilde{\mathbf{X}}_n$  représente un outil jusqu'ici inexploité de réduction de la dimension bi-directionnel (i.e., réduction simultanée sur la largeur et la longueur des images) qui pourrait être utilisé pour des tâches de classification ou de recherche rapide de similarité sur les données. Nous proposons donc d'étudier et de valider les performances du descripteur STPCA quant à réduire la dimension d'images.

## 4 Réduction d'images par l'approche STPCA

Les expérimentations proposées ont pour but de démontrer les capacités de la STPCA à fortement réduire la dimension d'images tout en conservant au mieux les relations entre elles. Les expériences concernant la réduction d'images ont été exercées à l'aide d'un processeur à 2.8 GHz utilisant 2.5 GiB de mémoire.

### 4.1 La base de données *Mnist*

La base de données *Mnist*<sup>1</sup> contient 70000 images représentant des chiffres écrits à la main. Cette base de données se décompose en deux parties : un ensemble d’entraînement de 60000 images et un ensemble de test contenant 10000 images. La taille de chaque image en pixels est de  $28 \times 28$ .



FIG. 1 – Exemple des 50 premières images de chiffres écrits à la main de *Mnist*.

La figure 1 illustre les 50 premières images de *Mnist*. Cette base de données d’images est également accompagnées d’un fichier d’étiquetage contenant le chiffre associé à chaque image. *Mnist* est ainsi considéré comme "une bonne base de données pour les gens qui cherchent à essayer des techniques d’apprentissage et de reconnaissance de motifs à partir de données réelles" (Y. LeCun, 1998, <http://yann.lecun.com/exdb/mnist/>).

### 4.2 Reconstruction d’image réduite par l’estimateur STPCA

Afin de donner un premier aperçu des performances de l’approche STPCA appliquée à des images, on s’intéresse dans un premier temps à l’estimateur STPCA  $\hat{X}_n$  de chaque image  $n$  de l’ensemble d’entraînement de *Mnist*. On applique donc une STPCA sur cet ensemble avec les paramètres d’entrées  $I = 28, J = 28$  et  $N = 60000$ . Les paramètres de réductions  $K$  et  $L$  sont déterminés de manière à ce que les matrices d’ordre réduit (i.e., les descripteurs STPCA) conservent la relations entre la longueur et la largeur des images d’origines. Dans le cas de *Mnist*, cela revient à considérer les valeurs telles que  $K = L$ .



FIG. 2 – Estimés STPCA d’images de *Mnist* obtenus à partir des matrices carrées d’ordre réduit de tailles  $1 \times 1$  à  $28 \times 28$ .

<sup>1</sup>La base de données *Mnist* est disponible à l’adresse web : <http://yann.lecun.com/exdb/mnist/>

La figure 2 montre les estimés STPCA des seize premières images de *Mnist* obtenus en fonction des valeurs des paramètres de réduction  $K$  et  $L$  avec  $K = L$ . Chaque ligne de cette figure correspond à une image et chaque colonne se réfère aux valeurs de  $K$  et  $L$ . La première colonne illustre les estimés STPCA calculés lorsque  $K = L = 1$ , la deuxième colonne les images obtenus lorsque  $K = L = 2$ , et ainsi de suite.

Cette première analyse graphique indique que de très faibles valeurs des paramètres de réduction, tels que les matrices d'ordre réduit sont de taille  $4 \times 4$  ou  $5 \times 5$  par exemple, correspondent à des estimés STPCA d'un degré de précision suffisant pour que l'oeil humain reconnaisse les chiffres qui y sont représentés.

### 4.3 Conservation des relations inter-images

En se focalisant sur le descripteur STPCA de chaque image  $n$  représenté par la matrice d'ordre réduit  $\tilde{\mathbf{X}}_n$ , on exploite entièrement le potentiel de réduction de la dimension de l'approche STPCA. Contrairement à l'estimateur STPCA, le descripteur STPCA n'est pas directement interprétable graphiquement car il contient des coefficients et non des valeurs associées à des pixels. Cependant, ces coefficients expriment les caractéristiques de chaque image et offre une nouvelle définition de celles-ci dans un espace réduit.

Les premières expériences consistent à mesurer l'impact de la réduction sur la conservation des relations de similarité entre les images. Pour cela, on applique l'approche STPCA à l'ensemble de test de *Mnist* puis on calcule les  $k$  plus proches voisins de chaque matrice des descripteurs STPCA  $\tilde{\mathbf{X}}_n$  (pour  $n = 1 \dots N$ ) parmi les  $N - 1$  autres matrices d'ordre réduit calculées. La distance entre deux matrices est obtenue par la norme de Frobenius de leur différence. La norme de Frobenius  $\|\mathbf{A}\|_F$  d'une matrice  $\mathbf{A}$  est donné par :

$$\|\mathbf{A}\|_F = \sqrt{\sum_{i=1}^I \sum_{j=1}^J (\mathbf{A}_{i,j})^2},$$

où  $\mathbf{A}_{i,j}$  correspond à la valeur de la ligne  $i$  à la colonne  $j$  de la matrice  $\mathbf{A}$ .

La qualité de la classification obtenue est ensuite mesurée par le taux de reconnaissance  $\tau_{rec}(n)$  de chaque image  $n$  qui indique la proportion de ses images voisines partageant la même étiquette qu'elle, i.e., la proportion de ses images voisines qui représentent le même chiffre que  $n$ . Le taux de reconnaissance d'une image  $n$  est donnée par :

$$\tau_{rec}(n) = \frac{|\mathcal{V}(n, E(n))|}{|\mathcal{V}(n)|}$$

où  $\mathcal{V}(n)$  est l'ensemble contenant les  $k$  plus proches voisins de  $n$ ,  $\mathcal{V}(n, E(n))$  est un sous-ensemble de  $\mathcal{V}(n)$  ne contenant que les voisins dont l'étiquette est égale à  $E(n)$ , l'étiquette de  $n$ , et  $|\mathcal{V}|$  symbolise le nombre d'éléments contenus dans l'ensemble  $\mathcal{V}$ .

La figure 3 illustre le taux de reconnaissance moyen obtenu à partir de l'ensemble d'entraînement de *Mnist* en fonction de la taille des matrices des descripteurs STPCA pour quatre différentes valeurs de  $k$ , le nombre de voisins pris en compte. Pour chaque valeur de  $k$ , le taux de reconnaissance obtenu à partir des images d'origine de *Mnist* (de taille  $28 \times 28$ ) est également représenté par une ligne en pointillés.

Ces tests démontrent la fiabilité du descripteur STPCA à fortement réduire la dimension tout en

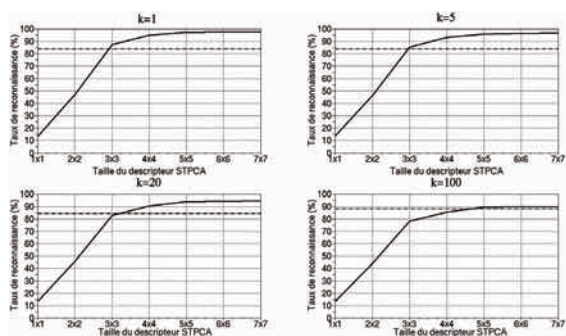


FIG. 3 – Taux de reconnaissance moyen en fonction de la taille du descripteur STPCA (matrice d'ordre réduit) pour différentes valeurs du nombre de voisins  $k$ .

conservant les relations entre les images. Pour les quatre valeurs de  $k$ , à partir du moment où les matrices d'ordre réduit sont au moins de taille  $3 \times 3$ , le taux de reconnaissance est supérieur ou égal à 80%. Grâce à la réduction du bruit consécutif à notre approche, lorsque  $k$  est très petit, le descripteur STPCA atteint même un taux de reconnaissance supérieur à celui obtenu avec les images d'origine à ce niveau de réduction. Lorsque  $k = 20$  et  $k = 100$ , ce comportement se produit également à partir du moment où les matrices d'ordre réduit sont respectivement de taille  $4 \times 4$  et  $5 \times 5$ .

#### 4.4 Apprentissage rapide de chiffres

La qualité des résultats précédents démontre l'efficacité de notre approche à fournir un bon descripteur de faible dimension pour chaque image. Ceci nous a conduit à considérer les capacités du descripteur STPCA à aider l'identification des chiffres représentés sur les images de l'ensemble de test de *Mnist* suite à une phase d'apprentissage exercée sur les images de l'ensemble d'entraînement de *Mnist*. La procédure suivie lors de cette expérience est la suivante.

Dans un premier temps, une STPCA est appliquée sur l'ensemble d'entraînement de *Mnist*. Ceci permet d'apprendre les matrices des modes propres  $\mathbf{P}$  et  $\mathbf{Q}$  ainsi que de calculer le descripteur STPCA de chaque image de l'ensemble d'entraînement. Ensuite, on s'intéresse à l'ensemble de test de *Mnist* dont on estime le descripteur STPCA de chaque image à partir des matrices  $\mathbf{P}$  et  $\mathbf{Q}$  apprises lors de l'étape précédente. La dernière étape consiste à calculer les  $k$  matrices d'ordre réduit de l'ensemble d'entraînement les plus proches de chaque descripteur STPCA des images de l'ensemble de test. Finalement, on affecte à chaque image de l'ensemble de test l'étiquette (i.e., le chiffre) le plus fréquemment représenté parmi ses voisins. La qualité de la classification obtenue est mesurée par le taux d'erreur de test proposé par LeCun et al. (1998) qui indique la proportion d'images dont le chiffre affecté ne correspond pas au chiffre affiché.

La figure 4 montre l'évolution du taux d'erreur de test en fonction de la taille du descripteur STPCA utilisé lors de l'apprentissage pour différentes valeurs de  $k$ , le nombre de voisins con-

## Réduction bi-directionnelle d'images

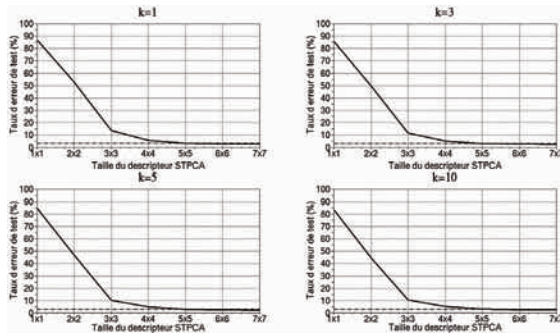


FIG. 4 – Taux d'erreur de test en fonction de la taille du descripteur STPCA (matrice d'ordre réduit) pour différentes valeurs du nombre de voisins  $k$ .

sidérés. Pour chaque valeur de  $k$ , le taux d'erreur de test obtenu avec les images originales de *Mnist* sans passer par le descripteur STPCA est également représenté par une ligne en pointillés. Une fois de plus, les résultats sont étonnamment bon car, quelque soit la valeur de  $k$ , les taux d'erreur de test sont inférieures à 5% à partir du moment où les matrices d'ordre réduit sont au moins de taille  $4 \times 4$ . De plus, pour toutes les valeurs de  $k$ , à partir du moment où les matrices d'ordre réduit sont de taille  $5 \times 5$  le taux d'erreur de test est plus petit que celui obtenu avec les images originales, atteignant des valeurs inférieures à 2.5%.

Avec descripteur STPCA	$1 \times 1$	$2 \times 2$	$3 \times 3$	$4 \times 4$	$5 \times 5$	$6 \times 6$	$7 \times 7$
	11	17	27	41	57	77	101
Avec images originales	1443						

TAB. 1 – Comparaison des temps de calcul moyen (en secondes) des  $k$  plus proches voisins (phase d'affectation) lorsqu'on utilise le descripteur STPCA ou les images originales.

Le tableau 1 illustre les temps de calcul moyen des  $k$  plus proches voisins, pour les quatre valeurs de  $k$ , lors de l'étape d'affectation des chiffres à la fin du processus d'apprentissage obtenus suite à l'utilisation ou non du descripteur STPCA. Dans les cas où on utilise le descripteur STPCA, les résultats sont exposés en fonction de la taille des matrices d'ordre réduit correspondantes.

En plus de fournir un apprentissage plus précis des chiffres de *Mnist*, le descripteur STPCA permet également de procéder à cet apprentissage dans un temps très inférieur à celui obtenu avec les images originales. Par exemple, le temps de calcul de la phase d'affectation est diminué de plus de 25 fois lorsqu'on utilise les descripteurs de taille  $5 \times 5$  plutôt que les images originales, tout en atteignant un taux d'erreur de test du même ordre, inférieur à 3%. Ces résultats sont encore plus surprenant quand on considère que le calcul des matrices des modes propres et des descripteurs STPCA se fait en moins de 60 secondes !



## 5 Relations aux méthodes existantes

Les expériences proposées pour valider notre approche ont démontré ses capacités à fortement réduire la dimension d'images tout en conservant suffisamment les relations pour efficacement procéder à des tâches de classification ou d'apprentissage à partir des versions réduites. Cette section situe les résultats obtenus dans le domaine de l'apprentissage d'images et présente les relations entre notre approche et les méthodes existantes de réduction de la dimension.

### 5.1 Apprentissage des images de *Mnist*

Depuis la publication de la base de données *Mnist*, plusieurs méthodes ont été proposées afin d'apprendre les chiffres écrits à la main représentés par les images. Les auteurs de LeCun et al. (1998) présentent un résumé des performances de plusieurs méthodes sur la base de données *Mnist* telles que les classifieurs linéaires, les  $k$  plus proches voisins, les Machines à Vecteurs Support virtuelles, les réseaux de neurones multi-niveaux, et les réseaux à convolution. D'autres méthodes ont également été proposées : la combinaison d'extracteurs de caractéristiques entraînés avec des Machines à Vecteurs Support dans Lauer et al. (2007), l'application des  $k$  plus proches voisins avec déformation non-linéaires dans Keysers et al. (2007), la combinaison de Machines à Vecteur Support et de l'apprentissage non supervisé de caractéristiques parcimonieuses dans Labusch et al. (2008), ou l'utilisation de réseaux à convolution large (Ranzato et al., 2006).

Afin de mesurer leurs performances, toutes ces méthodes utilisent le taux d'erreur de test proposé par LeCun et al. (1998), ce qui les rend directement comparables entre elles et avec notre approche. Selon ce critère, l'utilisation du descripteur STPCA et des  $k$  plus proches voisins n'est pas aussi performante que la meilleure méthode connue, à savoir les réseaux à convolution large de Ranzato et al. (2006), qui obtiennent un taux d'erreur de test de seulement 0.39%. Cependant, notre approche offre un apprentissage de meilleure qualité que plusieurs autres méthodes incluant tous les réseaux de neurones multi-niveaux présentés dans LeCun et al. (1998). En outre, il faut préciser que le descripteur STPCA n'est pas une méthode d'apprentissage mais une méthode de réduction de dimension. Lors de nos validations expérimentales, nous avons combiné notre approche à la recherche des  $k$  plus proches voisins, mais elle pourrait tout aussi aisément être combinée à des techniques plus complexes comme l'intégralité des méthodes existantes mentionnées précédemment. Il faut ainsi considérer le descripteur STPCA comme une technique de préparation de données antérieure à la phase d'apprentissage. Cette approche permettrait de réduire les temps de calculs et, suite aux résultats observés lors de la combinaison avec les plus proches voisins, elle permettrait probablement également d'améliorer les résultats obtenus lors de la phase d'apprentissage en comparaison de ceux calculés à partir des images originales. Remarquons également que, si une méthode d'apprentissage nécessite une phase de préparation sur les données (e.g., brouillage, sous-échantillonnage...), celle-ci peut toujours avoir lieu avant le calcul des descripteurs STPCA.

### 5.2 Méthodes d'extraction de caractéristiques

Tout comme les méthodes énoncées dans la section 2, le descripteur STPCA est une méthode d'extraction des caractéristiques. Cependant, son approche bi-directionnelle exploitant

simultanément la largeur et la longueur des images fait que la STPCA tient compte de la structure 2D des images lors de leur réduction. En outre, le descripteur STPCA peut très aisément être combiné aux autres techniques d'extraction de caractéristiques. On se dirigerait alors vers une extraction de caractéristiques bi-niveaux qui consisterait dans un premier temps à calculer le descripteur STPCA de chaque image, puis à appliquer une méthode d'extraction de caractéristiques standard (e.g., MDS, Isomap) à partir de ces descripteurs. Notons que, bien qu'étant fondée sur l'Analyse en Composantes Principales, la STPCA soit une méthode linéaire de réduction de la dimension, si le descripteur STPCA est combiné à une méthode non-linéaire de réduction de la dimension, comme Isomap ou LLE, la méthode bi-niveaux résultant de cette combinaison représente elle-même une méthode non-linéaire.

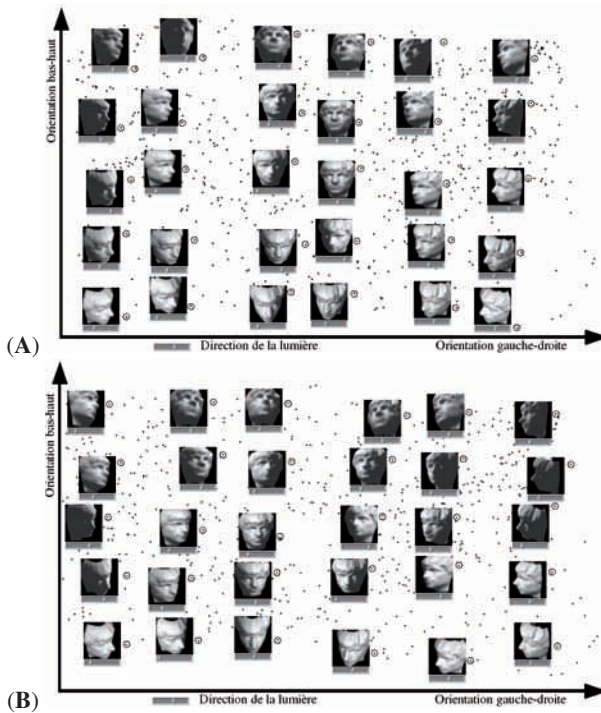


FIG. 5 – (A) Projection sur un hyperplan à trois dimensions des images originales de taille  $64 \times 64$  de faces par Isomap ( $k = 6$ ); (B) Projection sur un hyperplan à trois dimensions des descripteurs STPCA de taille  $12 \times 12$  des images de faces par Isomap ( $k = 7$ ).

Les figures 5A et 5B permettent de comparer les résultats obtenus lorsqu'on projette les images de la base de données *faces*<sup>2</sup> sur un hyperplan à trois dimensions. Cette base de don-

<sup>2</sup>La base de données *faces* est disponible à l'adresse web : <http://waldron.stanford.edu/~isomap/>

nées contient  $N = 698$  images de taille  $64 \times 64$  pixels ( $I = J = 64$ ) représentant un visage sous différentes postures et différents éclairages. Sur les figures 5A et 5B, chaque point de l'hyperplan représente une image de *faces* et les images associées à certains points, entourés sur la figure, sont illustrées à côté d'eux. Par soucis d'ergonomie, la troisième dimension n'est illustrée, par un curseur horizontal représenté sous les images, que pour les points des données clairement identifiés. La figure 5A reprend les résultats d'Isomap démontrés dans Tenenbaum et al. (2000) quant à son application sur *faces*. Sur cette figure, la valeur du paramètre  $k$  déterminant le nombre de voisins pris en compte par Isomap est fixé à  $k = 6$  comme proposé dans le papier. La figure 5B illustre les résultats obtenus lorsqu'on combine le descripteur STPCA et Isomap. Dans ce cas, le descripteur STPCA est calculé tel que les matrices d'ordre réduit sont de taille  $12 \times 12$  et Isomap est appliqué avec  $k = 7$ .

On peut remarquer que les résultats de l'approche d'extraction de caractéristiques bi-niveaux proposée sont d'au moins aussi bonne qualité que ceux obtenus avec l'utilisation exclusive d'Isomap. Les hyperplans calculés dans les deux cas sont très similaires et s'interprètent de la même manière : l'axe des abscisses est représentatif de la posture gauche-droite des visages sur les images, l'axe des ordonnées est représentatif de leur posture haut-bas et l'axe des hauteurs est représentatif de la direction de la lumière. Ceci démontre une fois de plus la fiabilité du descripteur STPCA à conserver les relations entre les images tout en réduisant la dimension. De plus, l'approche bi-niveaux s'avère intéressante car, dans le cas illustré, Isomap est appliqué à partir de matrice de taille  $12 \times 12$  au lieu de  $64 \times 64$  ce qui permet à la fois de réduire l'espace mémoire occupé ainsi que les temps de calculs lors du calcul de l'hyperplan. Finalement, suite aux résultats précédemment observés sur *Mnist*, on peut potentiellement espérer une meilleure qualité des hyperplans obtenus avec l'approche bi-niveaux qu'avec les méthodes conventionnelles.

## 6 Conclusion

Ce papier traite de l'utilisation du descripteur STPCA, une méthode spatio-temporelle d'extraction de caractéristiques dans le domaine de la perception visuelle. Les validations numériques effectuées démontrent les capacités de l'approche proposée à réduire la dimension tout en conservant au mieux les relations entre les images ; l'identification des chiffres écrits à la main des images de *Mnist* atteignant, en un temps réduit, un taux d'erreur plus faible à partir des descripteurs réduits des images qu'à partir de leur version originale. La combinaison de l'approche proposée à l'algorithme Isomap, une méthode non-linéaire de réduction de la dimension, fournit également des premiers résultats concluant, ouvrant vers une approche bi-niveaux d'extraction de caractéristiques.

Les travaux futurs concernent principalement la combinaison de l'approche proposée à des méthodes complexes d'apprentissages (Machines à Vecteur Support, réseaux de neurones à convolution...) ainsi qu'à d'autres méthodes, linéaires et non-linéaires, d'extraction des caractéristiques, qui pourrait mener à une approche innovante d'extraction de caractéristiques multi-niveaux.

## Références

- Bauzer-Medeiros, C., M. Joliveau, G. Jomier, et F. DeVuyst (2008). Managing sensor data on urban traffic. *Advances in Conceptual Modeling – Challenges and Opportunities 5232/2008*, 385–394.
- Joliveau, M. (2008). *Réduction de séries chronologiques de trafic routier urbain issues d'un réseau de capteurs géoréférencés et extraction de motifs spatio-temporels*. Ph. D. thesis, Ecole Centrale Des Arts Et Manufactures (Ecole Centrale Paris).
- Joliveau, M. et F. DeVuyst (2007). Space-time summarization of multisensor time series. case of missing data. In *Proc. of 2007 International Workshop on Spatial and Spatio-Temporal Data Mining*, pp. 631–636.
- Joliveau, M. et F. DeVuyst (2008). Recherche de motifs spatio-temporels de cas atypiques pour le trafic routier urbain. *Extraction et Gestion de Connaissances EGC 08, Revue des Nouvelles Technologies de l'Information - RNTI - E11, F. Guillet et B. Trousse 2*, 523–534.
- Jolliffe, I. (1986). *Principal component analysis*. New York : Springer-Verlag.
- Keysers, D., T. Deselaers, C. Gollan, et H. Ney (2007). Deformation models for image recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence 29*(8), 1422–1435.
- Kruskal, J. B. et M. Wish (1978). *Multidimensional Scaling*. Sage Publications.
- Labusch, K., E. Barth, et T. Martinetz (2008). Simple method for high-performance digit recognition based on sparse coding. *IEEE transactions on neural networks 19*, 1885–1889.
- Lauer, F., C. Suen, et G. Bloch (2007). A trainable feature extractor for handwritten digit recognition. *Pattern Recognition 40*, 1816–1824.
- LeCun, Y., L. Bottou, Y. Bengio, et P. Haffner (1998). Gradient-based learning applied to document recognition. In *Proc. of the IEEE*, Volume 86, pp. 2278–2324.
- Ranzato, M., C. Poultney, S. Chopra, et Y. LeCun (2006). Efficient learning of sparse representations with an energy-based model. In *Proc. of Advances in Neural Information Processing System NIPS 2006*.
- Richardson, M. (1938). Multidimensional psychophysics. *Psychological Bulletin 35*, 659–660.
- Roweis, T. et L. Saul (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science 290*, 2323–2326.
- Tenenbaum, J., V. de Silva, et J. Langford (2000). A global geometric framework for nonlinear dimensionality reduction. *Science 290*, 2319–2323.

## Summary

Inspired from human brain accuracy to recognize elements by seeing them, dimensionality in the domain of visual perception consists in extracting a small number of relevant perceptual features of an image data base in order to identify them.

This paper presents an innovative bi-directionnal approach partially based on a spatio-temporal method to accurately extract images features. Numerical experiments performed on 70000 handwritten digits images and on 698 images displaying a face demonstrate the accuracy of our approach to strongly reduce dimension while conserving intelligible relations between data objects, reaching more accurate classification from reduced version of images than from full-size ones.