



HAL
open science

Méthodes de reconnaissance de formes basées sur la carte de dissimilarité locale pour la classification des images

Youssef Ech-Choudany, Frederic Morain-Nicolier, Jérôme Landré, Benaïssa Bellach, Mustapha Assarar, Daniel Scida

► To cite this version:

Youssef Ech-Choudany, Frederic Morain-Nicolier, Jérôme Landré, Benaïssa Bellach, Mustapha Assarar, et al.. Méthodes de reconnaissance de formes basées sur la carte de dissimilarité locale pour la classification des images. GRETSI 2017, Sep 2017, Juan-Les-Pins, France. hal-03140040

HAL Id: hal-03140040

<https://hal.science/hal-03140040v1>

Submitted on 12 Feb 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Méthodes de reconnaissance de formes basées sur la carte de dissimilarité locale pour la classification des images

Youssef ECH-CHOUDANY^{1,2}, Frédéric MORAIN-NICOLIER¹, Jérôme LANDRÉ¹, Benaïssa BELLACH², Mustapha ASSARAR³, Daniel SCIDA³

¹CReSTIC, URCA-IUT de Troyes 9 rue de Québec CS 90396, 10026 TROYES Cedex, France

²LSE2I, UMP-ENSA de Oujda Hay Qods BP 669 OUJDA, Maroc

³LISM, URCA-IUT de Troyes 9 rue de Québec CS 90396, 10026 TROYES Cedex, France

y.echchoudany@gmail.com

Résumé – La carte de dissimilarité locale (CDL) est une mesure pour comparer localement deux images binaires ou en niveaux de gris. Ce travail présente une adaptation de la CDL comme mesure de similarité pour la reconnaissance de formes. Dans le cadre d'un problème de classification, des expériences sur la base de chiffres manuscrits MNIST ont été réalisées. La comparaison entre les performances d'un réseau de neurones artificiels (ANN), les k -plus proches voisins (k NN) basés sur la CDL (ANN-CDL, k NN-CDL), la distance euclidienne (ANN-Euclidienne, k NN-Euclidienne) et la similarité structurale moyenne (ANN-MSSIM, k NN-MSSIM) permet de mettre en évidence les bonnes capacités discriminatives de la CDL. L'apprentissage de ANN et k NN par des bases de données de volumes différents a montré que les méthodes ANN-CDL et k NN-CDL convergent plus rapidement par rapport aux autres méthodes. Une précision de 97,6% a été obtenue par la méthode ANN-CDL.

Abstract – The local dissimilarity map (LDM) is a measure to compare locally two binary or grayscale images. This work presents an adaptation of the LDM as a similarity measure for pattern recognition. Experiments on MNIST digit database were made. The comparison between the performances of an artificial neural network (ANN), k -nearest neighbors based LDM (ANN-LDM, k NN-LDM), Euclidean distance (ANN-Euclidean, k NN-Euclidean), and mean structural similarity (ANN-MSSIM, k NN-MSSIM) have proved the best discriminative capacities of the LDM. ANN and k NN classifiers have been performed using various learning datasets of MINIST database and show that the ANN-LDM and k NN-LDM methods converge more rapidly compared to the other methods. An accuracy of 97.6% was obtained by the ANN-LDM method.

1 Introduction

De nos jours, de nombreuses techniques d'analyse et de traitement ont été étudiées pour classifier les images [1, 2]. D'une manière générale, le principe de ces méthodes peut être divisé en deux étapes, à savoir l'extraction de caractéristiques et la classification. La première étape vise à extraire des vecteurs de descripteurs à partir des caractéristiques distinctives locales des images. Dans la deuxième étape, un classifieur supervisé est formé sur tous les vecteurs résultants pour reconnaître la classe des images inconnues. L'exactitude de la classification dépend de la qualité des descripteurs.

La reconnaissance des chiffres manuscrits [3, 4] est une branche importante de la reconnaissance d'image dans le domaine de la reconnaissance des formes. L'objectif de ce travail, est de proposer une méthode de reconnaissance de formes basée sur une mesure de similarité entre les images. L'avantage de cette méthode est de ne pas nécessiter d'extraction de descripteurs.

De nombreuses mesures de similarités ont été utilisées dans le domaine du traitement des images (comparaison, débruitage, compression, qualité, etc). Parmi elles, on trouve la carte de dissimilarité locale (CDL) [5]. Cette méthode permet une évalua-

tion bien définie des dissimilarités locales entre deux images, ce qui n'est pas le cas des autres méthodes. Cette mesure a été développée dans un premier temps pour comparer les images binaires, ensuite une extension de la méthode a été adaptée pour les images en niveaux de gris. La CDL a montré de bonnes performances par rapport aux autres méthodes classiques de comparaison d'images. Notre objectif est d'exploiter cette comparaison comme distance entre les images.

Dans un premier temps, le principe des méthodes de classification basées sur une mesure de similarité est introduit. Dans un second temps, le principe de la CDL pour la comparaison des images en niveaux de gris est présenté. La partie résultats est consacrée à une comparaison de notre proposition et l'approche de l'état de l'art. Les comparaisons entre les performances d'un réseau de neurones artificiels (ANN) et la méthode des k -plus proches voisins (k NN) basées sur la distance euclidienne (ANN-Euclidienne, k NN-Euclidienne), sur la similarité structurale moyenne (SSIM) [6] (ANN-MSSIM, k NN-MSSIM) et sur la CDL (ANN-CDL, k NN-CDL), sont présentées.

2 Méthodes de classification basées sur une mesure de similarité

Plusieurs méthodes de reconnaissance de formes pour la classification des images ont été développées. La plupart de ces méthodes sont basées sur l'approche de l'extraction de descripteurs. La limite de cette approche est liée à la spécificité des descripteurs utilisés au problème étudié. Dans cette section, une approche basée sur les mesures de similarités entre les objets est introduite. Les classifieurs basés sur cette approche [7] estiment le libellé de la classe d'un échantillon d'essai en se basant sur les similarités entre l'échantillon d'essai et l'ensemble des échantillons d'apprentissage.

Dans l'approche usuelle, les objets sont représentés dans un espace de descripteurs. D'autres manières peuvent être trouvées en construisant des règles de décision basées sur des mesures de similarités entre les objets. De telles mesures deviennent efficaces lorsque les objets à classer sont constitués d'un grand ensemble d'attributs (les images) et les descripteurs ne peuvent pas les caractériser.

Lorsqu'une image est présentée pour être classée, elle est comparée aux autres images et le résultat de cette comparaison ne peut être que «similaire» ou «dissimilaire». Pour mesurer le degré de similarité entre deux images, on considère la fonction d , définie par : $d : E \times E \rightarrow [0, \infty]$. En classification, d doit être une métrique ou semi-métrique.

En effet, plus la mesure d entre deux images tend vers zéro plus les deux images sont similaires. Ces mesures peuvent donc être traitées comme une description d'un espace de caractéristique spécifique, où chaque dimension correspond à un objet. Cela ne modifie pas essentiellement l'approche basée sur les descripteurs. En général, tout classifieur supervisé fonctionnant sur des descripteurs peut être ainsi utilisé. Dans le processus d'apprentissage (Fig. 1), les classifieurs sont construits sur la matrice de distance D d'une taille $m \times m$ (où chaque valeur de donnée exprime une mesure de similarité entre deux objets d'apprentissage) : $D(k, l) = d(I_k, I_l)$, d est la distance entre les deux objets I_k et I_l . Les p objets de test sont classés en utilisant leurs distances aux m échantillons d'apprentissage (la taille des données d'essai est $(p \times m)$). Chaque image est ainsi représentée par l'ensemble de ses distances aux autres images. Le travail du classifieur est donc de déterminer les seuils permettant de distinguer les images similaires des images dissimilaires.

3 Carte de dissimilarité locale

3.1 CDL pour image binaire

Nous rappelons ici brièvement le principe de la carte de dissimilarité locale entre deux images binaires [5]. L'objectif est de comparer localement deux images binaires avec une mesure locale bien définie. La carte de dissimilarité locale regroupe l'ensemble des mesures de dissimilarités locales faites

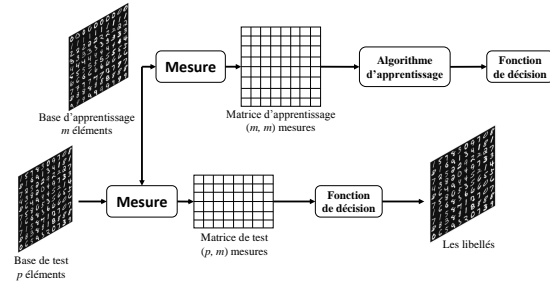


FIGURE 1 – Processus des méthodes de classification basées sur une mesure de similarité.

en différents endroits des images. Ceci permet d'avoir accès à leur répartition spatiale. Ces informations sont rarement accessibles via les méthodes de comparaison d'images existantes qui sont souvent globales. Les images binaires étant représentées comme des ensembles de points, une distance locale efficace est une version modifiée de la distance de Hausdorff. La définition suivante de la CDL permet d'obtenir ses valeurs avec un coût de calcul faible :

Définition 1. CDL binaire

Soient A et B deux images binaires, la CDL est définie de $\mathbb{R}^2 \times \mathbb{R}^2$ dans \mathbb{R}^2 par

$$CDL(A, B)(p) = |A(p) - B(p)| \max(d(p, A), d(p, B)), \quad (1)$$

$p = (x, y)$ étant une coordonnée pixel et $d(p, A)$ la transformée en distance de A c'est-à-dire la distance entre le pixel p et le pixel de A non-nul (donc définissant la forme) le plus proche.

3.2 CDL pour image en niveaux de gris

Plusieurs approches sont possibles pour généraliser la CDL aux images en niveaux de gris. Une solution consiste à étendre la définition de la transformée en distance aux images en niveaux de gris. La plupart des propositions présentent l'inconvénient de mêler distances pures et écarts de luminance. Les distances obtenues ne sont plus des écarts de position entre deux pixels (x, y) . La proposition de Molchanov se distingue par la définition d'une distance résultant de l'accumulation de distances entre les sous-images binaires successives des deux images en niveaux de gris (Real Valued Distance Transform (RVDT)) [8] :

$$RVDT(p, A) = \frac{1}{b-a} \sum_{i=a}^b d(p, A_i) \omega_i \quad (2)$$

avec $A_i = \{p : A(p) \geq i\}$, $0 \leq a < b \leq 255$ et $\omega_i = 1$ dans le cas le plus simple. Une possibilité serait donc d'exploiter cette RVDT en l'injectant dans l'éq. (1). Néanmoins, des distances entre niveaux de gris et entre positions (x, y) sont alors mélangées, rendant délicate une éventuelle interprétation et posant le problème des dynamiques des dimensions.

Nous proposons ici de conserver l'esprit de l'approche de Molchanov en découpant les images en niveaux de gris en images binaires (avec des seuils donnés). Les images binaires peuvent

être comparées avec la CDL de l'eq. (1). Les CDL sont alors accumulées en une moyenne selon les seuils.

Définition 2. Proposition de CDL pour image en niveaux de gris

Soient A et B deux images en niveaux de gris, N est un entier représentant le nombre de coupes, la CDL_N est définie de $\mathbb{R}^2 \times \mathbb{R}^2$ dans \mathbb{R}^2 par :

$$CDL_N(A, B)(p) = \frac{1}{N} \sum_{i=1}^N CDL(A_i, B_i)(p), \quad (3)$$

où A_i (resp. B_i) est l'image binaire de A (resp. B) pour la i -ème coupe donnée par la sous-image de A :

$$A_i = \{p : A(p) \geq s_i\}. \quad (4)$$

Parmi les multiples stratégies pour déterminer les seuils donnant les sous-images, nous choisissons ici une solution simple qui consiste à prendre des seuils régulièrement espacés entre 0 et le maximum des deux images. Ce qui est équivalent à une normalisation en luminosité des deux images :

$$s_i = \frac{i}{N} m_{AB}, i \in [1..N] \quad (5)$$

avec $m_{AB} = \max(\max_p A(p), \max_p B(p))$.

La CDL finale entre deux images en niveaux de gris est obtenue en augmentant le nombre de coupes :

$$CDL(A, B)(p) = \lim_{N \rightarrow \infty} CDL_N(A, B)(p). \quad (6)$$

Cette définition généralise complètement la définition de la CDL binaire. Si les images sont binaires, toutes les coupes sont identiques et l'équation (6) revient à l'équation (1).

En pratique, si les niveaux de gris sont codés sur 8 bits, on peut poser $N = 255$.

La CDL compare entre les structures des images. Lorsque deux images ont la même structure mais différentes luminosités, leur CDL sera nulle. Donc la luminosité moyenne n'a aucune influence sur la comparaison des images par la CDL.

4 Base de données

La base de données MNIST (Fig. 2) contient un grand nombre de chiffres manuscrits, et se compose de 10 classes (chiffres de 0 à 9). Chaque chiffre est une image de 28x28 pixels. La base MNIST est disponible à partir de la page web <http://yann.lecun.com/exdb/mnist>, où de nombreux résultats de classification de différentes méthodes sont également répertoriés. Les images sont des images en niveaux de gris. La base dispose d'un ensemble d'apprentissage de 60 000 exemples et d'un ensemble de tests de 10 000 exemples.

5 Résultats et discussion

Pour évaluer les performances de la CDL dans un algorithme d'apprentissage supervisé, la base de données MNIST est utilisée comme application. Les méthodes ANN et k NN sont exploitées comme algorithmes d'apprentissage. Le ANN utilisé

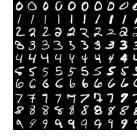


FIGURE 2 – Exemple de chiffres manuscrits de la base de données MNIST.

est de type 'Feedforward Network' avec 10 couches cachées. Le nombre de voisins considéré pour le k NN est $k=1$. Ces paramètres sont ceux qui nous a permis d'obtenir les meilleures précisions. Les performances de la CDL ont été comparées avec celles de la distance euclidienne et de la MSSIM. Les trois distances sont définies par :

$$d_{CDL}(A, B) = \|CDL(A, B)\|_1 \quad (7)$$

$$d_{Euclidienne}(A, B) = \|A - B\|_1 \quad (8)$$

$$d_{MSSIM}(A, B) = 1 - MSSIM(A, B) \quad (9)$$

Pour mesurer les performances des trois distances utilisées, la précision a été évaluée sur huit bases d'apprentissage (61, 121, 241, 481, 801, 2401, 6003 et 12000 individus) et une base de test (2003 individus).

Les figures 3 et 4 présentent successivement, les résultats obtenus par le ANN et le k NN. Ces figures montrent les précisions obtenues par les trois distances (CDL, Euclidienne et MSSIM) en fonction du nombre d'individus de la base d'apprentissage. On note que les méthodes basées sur la CDL atteignent une précision de 95% pour une base d'apprentissage de 2401 individus, par contre les autres méthodes atteignent la même précision pour une base d'apprentissage de 12005. Ceci signifie que les méthodes ANN-CDL et k NN-CDL convergent plus rapidement vers leurs précisions maximales par rapport aux autres méthodes.

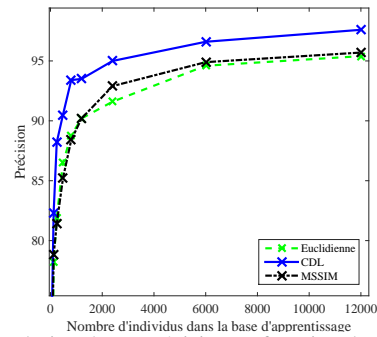


FIGURE 3 – Évolution de La précision en fonction du nombre d'individus de la base de données d'apprentissage. Méthode de classification : ANN.

La figure 5 présente les résultats obtenus par les méthodes de k NN-CDL et de ANN-CDL. Il apparaît que les deux méthodes convergent au même temps vers leurs précisions maximales (96,80% et 97,60%). Par contre, la méthode ANN-CDL a la plus grande précision (97,60%).

Les résultats obtenus permettent de vérifier les capacités discriminatives de la CDL. L'avantage de la CDL par rapport aux

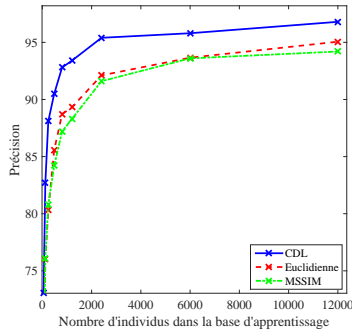


FIGURE 4 – Évolution de La précision en fonction du nombre d'individus de la base de données d'apprentissage. Méthode de classification : k NN.

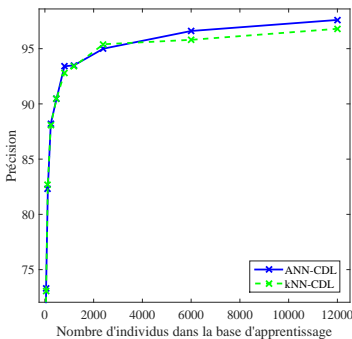


FIGURE 5 – Évolution de La précision en fonction du nombre d'individus de la base de données d'apprentissage pour les deux méthodes de classification (k NN-CDL et ANN-CDL).

autres distances est sa capacité à converger rapidement vers sa précision maximale. Cette rapidité permet à la CDL d'obtenir de bons résultats de classification même pour de faibles bases d'apprentissage. Cet avantage permettra à la CDL de résoudre des problèmes de classification dans des domaines où il y a très peu d'exemples d'apprentissage, comme le domaine de la médecine.

Les méthodes de classifications basées sur une mesure de similarité permettent d'exploiter la totalité de l'information présente dans les images. Ces méthodes permettent aussi d'éliminer la tâche d'extraction de caractéristiques sur laquelle les méthodes classiques sont basées.

On note que différentes méthodes ont été exploitées pour la classification de la base de données MNIST. La meilleure précision a été de l'ordre de 99,8% pour la base d'apprentissage complète. On note aussi que ce travail a deux principaux objectifs. Le premier est la présentation d'une méthode de classification sans extraction de caractéristiques. Le deuxième est la présentation d'une mesure de similarité qui permet de résoudre des problèmes de classification pour des bases d'apprentissage qui contiennent très peu d'éléments.

L'inconvénient de l'approche basée sur les mesures de similarités, dans des applications particulières, est la grande dimension de l'espace engendré par les similarités. Trois types de solutions peuvent être proposés pour résoudre ce problème. Le premier consiste à construire l'espace de similarité à l'aide

de tous les objets disponibles, ensuite appliquer certaines techniques de réduction de la dimensionnalité. Le but de la seconde est de sélectionner les descripteurs les plus pertinents en utilisant une technique de sélection. Le troisième type de solution fonctionne en choisissant directement un petit ensemble d'objets.

6 Conclusion

L'objectif de ce travail a été la mise en place d'une méthode de classification basée sur la CDL et sans exploitation de descripteurs. Deux méthodes de classification ont été exploitées pour mesurer les capacités discriminatives de la CDL (CDL-ANN, k NN-CDL). Les performances en terme de classification de la CDL ont été comparées avec celles de la distance euclidienne et la similarité structurale moyenne (MSSIM). Les résultats obtenus montrent que les méthodes de classification basées sur la CDL convergent plus rapidement avec une précision de 97,6% pour la méthode ANN-CDL.

Références

- [1] J. Zhang, M. Marszalek, S. Lazebnik, C. Schmid. *Local Features and Kernels for Classification of Texture and Object Categories : A Comprehensive Study*. International Journal of Computer Vision, 2007, Vol. 73, Issue 2, p. 213-238.
- [2] M. Luo, K. Zhang. *A hybrid approach combining extreme learning machine and sparse representation for image classification*. Engineering Applications of Artificial Intelligence, 2014, vol. 27, p. 228-235.
- [3] X X. Niu, C Y. Suen. *A novel hybrid CNNCSVM classifier for recognizing handwritten digits*. Pattern Recognition, 2012, vol. 45, no 4, p. 1318-1325.
- [4] L. Deng. *The MNIST database of handwritten digit images for machine learning research [best of the web]*. IEEE Signal Processing Magazine, 2012, vol. 29, no 6, p. 141-142.
- [5] É. Baudrier, F. Morain-Nicolier, G. Millon, S. Ruan. *Binary-image comparison with local-dissimilarity quantification*. Pattern Recognition, 2008 vol. 41, p. 1461-1478.
- [6] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli. *Image quality assessment : from error visibility to structural similarity*. IEEE transactions on image processing, 2004, vol. 13, no 4, p. 600-612.
- [7] Y. Chen, E. K. Garcia, M. R. Gupta, A. Rahimi, L. Cazzanti. *Similarity-based classification : Concepts and algorithms*. Journal of Machine Learning Research, 2009, vol. 10, p. 747-776.
- [8] I. Molchanov and P. Teran, *Distance transforms for real-valued functions*, Journal of Mathematical Analysis and Applications, 278, no. 2, p. . 472-484, 2003.