

INF8770

Technologies multimédias

H2024 - Travail pratique #3

Indexation de contenu pictural

Remise du travail :

- Au plus tard, le 8 avril, 9h sur Moodle - **aucun retard accepté**

Documents à remettre :

- Moodle : votre code source et un rapport **PDF** dans une archive zippée. Ne remettez pas les données.
- Turnitin : vous devez remettre votre rapport sur www.turnitin.com. Les instructions sont sur Moodle.

Autres directives :

- Il vous est interdit d'utiliser du code écrit par d'autres équipes.
- Il vous est permis d'utiliser du code trouvé sur internet et des librairies externes, mais vous devez nous donner toutes les références.
- Le code lui-même n'est pas évalué, mais il doit être remis au complet.

Pénalités éventuelles :

- Référence non citée (**-0.5 point**);
- Non remise du code ou du rapport sur Turnitin (**-0.5 point**);
- Le rapport n'est pas sous le format pdf (**-0.5 point**);
- Trop de chiffres significatifs dans les tableaux/figures (**-0.5 point**);
- Le document remis sur Moodle ne correspond pas exactement à celui remis sur Turnitin (**-0.5 point**)

Préambule : Ce travail pratique inclut une évaluation des qualités 2, 3 et 5 du Bureau canadien d'agrément des programmes de génie (BCAPG).

- **Qualité 2 Analyse de problèmes** : capacité d'utiliser les connaissances et les principes appropriés pour identifier, formuler, analyser et résoudre des problèmes d'ingénierie complexes et en arriver à des conclusions étayées.
- **Qualité 3 Investigation** : capacité d'étudier des problèmes complexes au moyen de méthodes mettant en jeu la réalisation d'expériences, l'analyse et l'interprétation des données et la synthèse de l'information afin de formuler des conclusions valides.
- **Qualité 5 Utilisation d'outils d'ingénierie** : capacité de créer et de sélectionner des techniques, des ressources et des outils d'ingénierie modernes et de les appliquer, de les adapter et de les étendre à un éventail d'activités simples ou complexes, tout en comprenant les contraintes connexes.

Travail à réaliser : Dans ce TP, vous devez comparer deux algorithmes de recherche de vidéos : à partir d'une image et d'une banque de données, vous devez retrouver la vidéo dans laquelle cette image est extraite. Les deux algorithmes reposent sur les descripteurs à base d'histogramme de couleur et de réseaux de neurones.

Objectif de l'algorithme de recherche de vidéos

Étant donnée une image JPEG, vous devez retrouver la vidéo MP4 dans laquelle cette image est extraite. Si vous trouvez une vidéo candidate, vous devez également retourner le moment où cette trame apparaît (en seconde). Si au contraire, aucune vidéo ne semble être candidate, aucun moment n'est retourné.

Description des données

La banque de vidéos contient 100 vidéos en HD (1280×720), sous le format MP4, libres de droit, extraites du site www.pexels.com (ces vidéos sont disponibles sur Moodle). En plus de ces vidéos, vous trouverez 1000 images JPEG qui seront les requêtes pour la recherche ainsi qu'un fichier `gt.csv` qui indique pour chaque image le nom de la vidéo dans laquelle elle est extraite et son minutage en secondes (dans le cas où l'image provient d'une vidéo absente de la banque de données, ces deux valeurs sont respectivement à `out` et `None`). Les images sous le format JPEG ont été extraite avec perte des vidéos.

Description des algorithmes

La figure 1 illustre l'approche générale de l'algorithme :

1. une banque de données est constituée de 100 vidéos MP4 ;
2. *phase d'indexation* : l'ensemble des trames des vidéos ou un sous-ensemble de taille N est encodé à l'aide d'un descripteur (chaque trame est encodé par un vecteur de dimension D) ;
3. *phase de recherche* : étant donnée une image JPEG, elle est d'abord encodée avec le même descripteur. Puis, à l'aide d'une mesure d'affinité (distance euclidienne, produit scalaire, similarité cosinus, etc), on retrouve dans la matrice la ligne (c'est-à-dire la trame) "la plus ressemblante". Si la similarité est assez forte, on retourne le nom de la vidéo et le minutage correspondant, sinon, on retourne (`out`, `None`).

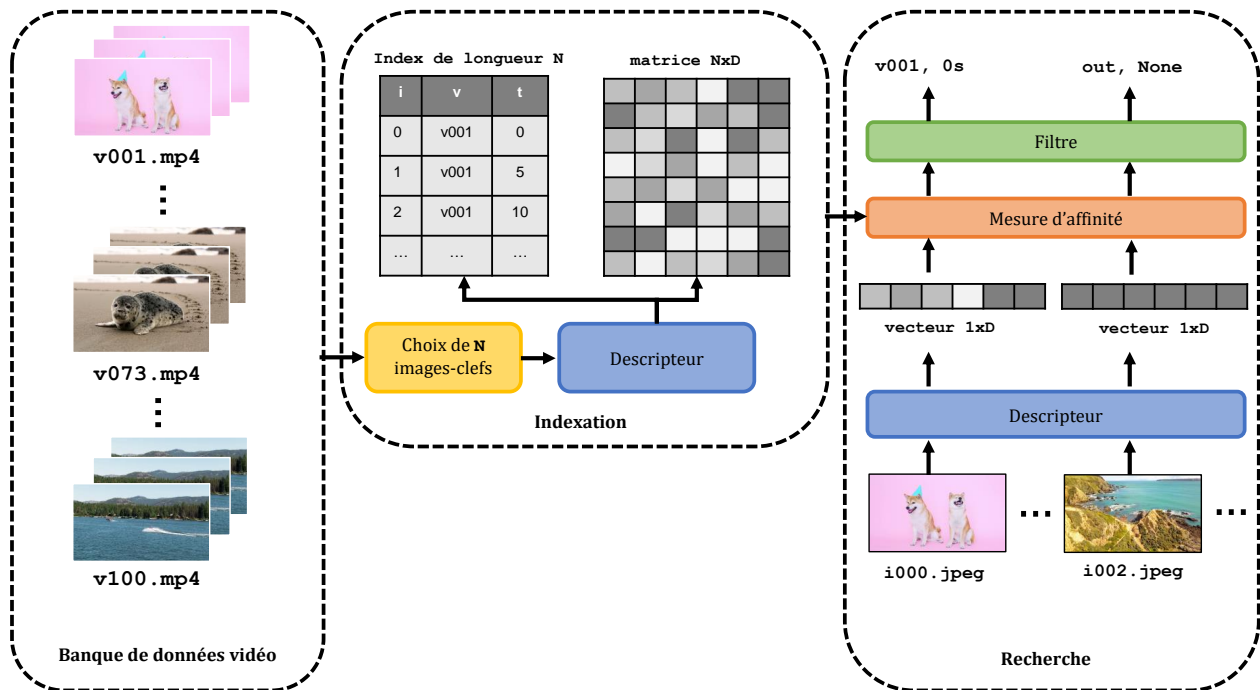


FIGURE 1 : Illustration de l'algorithme de recherche de vidéos

Format du fichier `gt.csv`

Le fichier contient trois colonnes :

- `image` : le nom de l'image ;
- `video` : le nom de la vidéo si elle est présente dans le dossier des vidéos, sinon `out` ;
- `minutage` : l'instant d'apparition de la trame dans la vidéo (en seconde) si cette vidéo est présente, sinon vide. La première trame de chaque vidéo correspond au minutage 0 seconde.

Format de votre fichier `test.csv`

Le fichier devra contenir trois colonnes :

- `image` : le nom de l'image ;
- `video_pred` : le nom de la vidéo candidate ou `out` ;
- `minutage_pred` : l'instant d'apparition de la trame candidate dans la vidéo (en seconde) si cette vidéo est présente, sinon vide. La première trame de chaque vidéo correspond au minutage 0 seconde.

Mesures de performance

Les cinq mesures de performance sont les suivantes :

1. le pourcentage de vidéos correctement retrouvées (en %) ;
2. l'écart (en valeur absolue) moyen sur le minutage lorsque la vidéo est correctement retrouvée et présente dans la banque de données (en seconde) ;
3. le taux de compression des données indexées, calculé avec $1 - \frac{T_c}{T_o}$ avec T_o la valeur obtenue à la question 1b) et T_c la taille en octet de la matrice de dimension $N \times D$;
4. le temps d'indexation des trames (en minute/seconde) ;
5. le temps de recherche d'une image (en milliseconde).

Pour mesurer les deux premiers mesures de performance, enregistrez vos résultats dans un fichier `test.csv` puis lancez la commande suivante :

```
python evaluate.py --file=../results/test.csv --file_gt=../data/gt.csv
```

Remarques :

- en termes de temps d'exécution, nous vous conseillons d'optimiser votre algorithme si vous constatez qu'il est lent (l'indexation devrait prendre moins de 5 minutes et la recherche moins de 100 ms par image). Lors du mesure du temps d'indexation et de recherche, ne comptez pas le temps de chargement des données et/ou modèles ;
- les requêtes sur les images JPEG ne doivent pas être parallélisées.

Question 1 (/4)

- a) Quelles seraient les performances, en termes de pourcentage de vidéos correctement retrouvées, d'un algorithme consistant à retourner la trame d'une vidéo dès que la trame est identique à une image JPEG ? Cet algorithme retournerait (`out`, `None`) si aucune trame n'est identique à l'image JPEG.
- b) Quelle est la taille en octet des 100 vidéos MP4 sur le disque dur ?
- c) Quel serait l'ordre de grandeur de la taille en octet si on extrayait toutes les trames des 100 vidéos en JPEG ?
- d) Quel serait l'ordre de grandeur de la taille en octet si on conservait en mémoire RAM toutes les trames non compressées (sur 8 bits) ?
- e) Comparez les trois tailles des questions b), c) et d) et expliquez l'origine des différences des ordres de grandeurs.

- **Qualité évaluée : 2.1 Identifier et formuler un problème**

Critère d'évaluation : Analyse de l'algorithme proposé en a).

Question 2 (/2)

Formulez et justifiez deux hypothèses sur l'algorithme de recherche de la figure 1 avec des descripteurs basés sur des histogrammes de couleurs par rapport à ces paramètres :

- le choix des N images-clefs : toutes, une sélection dynamique des images-clefs, quelques images-clefs par vidéo, etc ;
- les paramètres des histogrammes de couleurs : l'espace colorimétrique (RGB ou YUV), la dimensionalité (1D ou 3D) et la taille des intervalles d'histogramme ;
- la mesure d'affinité : la distance euclidienne, la distance de Bhattacharyya, la similarité cosinus, le produit scalaire, etc ;

- **Qualité évaluée : 3.1 Formuler des hypothèses testables**

Critère d'évaluation : Qualité et exhaustivité des hypothèses. Les hypothèses doivent être bien expliquées par rapport aux mesures de performance.

Question 3 (/3)

Implémentez l'algorithme en suivant les deux hypothèses formulées précédemment. Donnez les résultats de votre algorithme pour les cinq mesures de performance à chaque hypothèse.

- **Qualité évaluée : 2.4 Produire des résultats**

Critère d'évaluation : Pertinence des résultats et du protocole pour démontrer la qualité de la solution proposée.

- **Qualité évaluée : 5.3 Créer ou adapter un outil**

Critère d'évaluation : Capacité à créer un code informatique fonctionnel.

- **Qualité évaluée : 5.4 Intégrer des outils**

Critère d'évaluation : Capacité à bien intégrer les fonctions de bibliothèques de traitement d'images/vidéos dans son propre code.

Question 4 (/2)

Formulez et justifiez deux hypothèses sur l'algorithme de recherche de la figure 1 avec des descripteurs basés sur des réseaux de neurones par rapport à ces paramètres :

- le choix des N images-clefs : toutes, une sélection dynamique des images-clefs, quelques images-clefs par vidéo, etc ;
- les paramètres des réseaux de neurones : l'architecture du réseau (AlexNet, ResNet-18, ResNet-50, VGG-16, ViT, etc), les poids pré-entraînés du modèle, la couche de représentation (la dernière ou l'avant-dernière) et le composant informatique utilisé (CPU ou GPU) ;
- la mesure d'affinité : la distance euclidienne, la similarité cosinus, le produit scalaire, etc ;
- **Qualité évaluée : 3.1 Formuler des hypothèses testables**
Critère d'évaluation : Qualité et exhaustivité des hypothèses. Les hypothèses doivent être bien expliquées par rapport aux mesures de performance.

Question 5 (/3)

Implémentez l'algorithme en suivant les deux hypothèses formulées précédemment. Donnez les résultats de votre algorithme pour les cinq mesures de performance à chaque hypothèse.

- **Qualité évaluée : 2.4 Produire des résultats**
Critère d'évaluation : Pertinence des résultats et du protocole pour démontrer la qualité de la solution proposée.
- **Qualité évaluée : 5.3 Créer ou adapter un outil**
Critère d'évaluation : Capacité à créer un code informatique fonctionnel.
- **Qualité évaluée : 5.4 Intégrer des outils**
Critère d'évaluation : Capacité à bien intégrer les fonctions de bibliothèques de traitement d'images/vidéos dans son propre code.

Question 6 (/6)

a) Analysez les résultats obtenus selon les cinq mesures de performance et mettez-les en relation avec les hypothèses. Est-ce que les hypothèses sont supportées par les résultats ?

b) En prenant la meilleure configuration pour les descripteurs à base d'histogramme et à base de réseau de neurones, donnez un exemple d'image JPEG où seul l'algorithme à base d'histogramme de couleur se trompe sur la vidéo, un autre exemple où seul l'algorithme à base de réseaux de neurones se trompe sur la vidéo, et un dernier où les deux algorithmes de recherche se trompent sur la vidéo. Discutez.

- **Qualité évaluée : 3.5 Analyser les résultats expérimentaux**
Critère d'évaluation : Qualité et exhaustivité de l'analyse des résultats concernant leur cohérence avec le fonctionnement des méthodes appliquées.
- **Qualité évaluée : 3.6 Vérifier les hypothèses et argumenter**
Critère d'évaluation : Qualité et exhaustivité de l'analyse critique des résultats en fonction des hypothèses.