

Rapport Projet 2

Lucas Chabeau - Matthieu François - Etienne Hamard

22 avril 2019

Contents

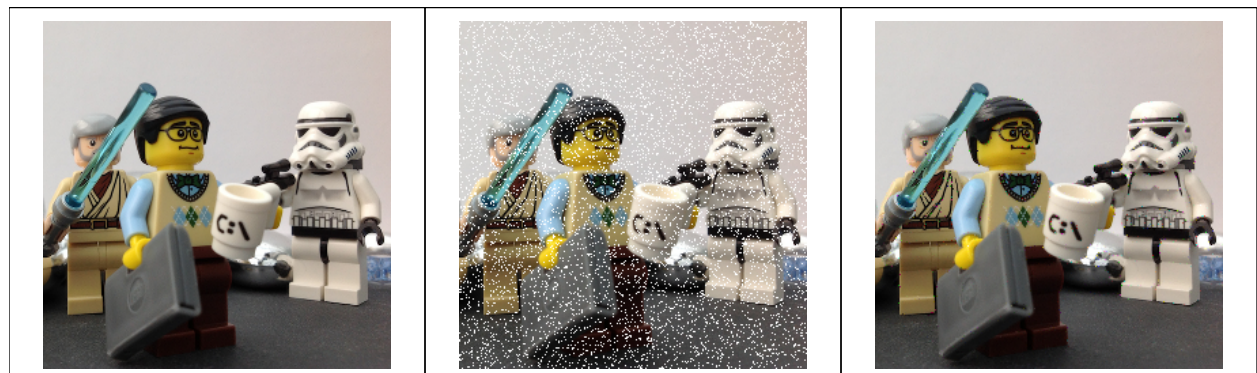
Introduction	1
1. Etude de l'image avec 10% de bruit	1
4. Impact du bruit sur l'efficacité de K	5
Conclusion	6
Annexe	6

Introduction

Dans le cadre du cours d'Analyse de données 2, nous nous sommes penchés sur la réalisation d'un projet concernant la restauration d'image. Cela avec la méthode des k plus proches voisins (knn). L'étude met à disposition une image déclinée en 4 versions. Une version complète, une version avec 10% de valeur manquante, une avec 20% et enfin une avec 50%. Les images sont composées de 3 jeux de données, le pourcentage de vert, de bleu et de rouge. Afin de réaliser l'algorithme des knn nous avons choisis de modifier un peu ces jeux de données. Par image nous avons réalisé trois knn, un par couleur, afin d'avoir un résultats plus précis. Il a donc extraire des trois images, les valeurs R, G et B pour les transformer en trois jeux de données différents, comportant la coordonnée x, et y associée.

1. Etude de l'image avec 10% de bruit

Le choix le plus important dans l'algorithme des k plus proches voisins est le nombre de voisins K sur lesquels s'appuyer pour prédire la classe des individus concernés choisir. La présence de l'image originale nous permet donc de calculer la précision que peut avoir l'algorithme pour chaque K. Pour cela nous avons utilisé le Mean Squared Error (MSE), étant donné qu'il n'est pas possible de comparer combien de pixels sont similaires à l'originale vu que nous avons divisé la restauration de ceux-ci en 3 (1 par couleur).



A première vue la restauration est efficace, l'image est proche de l'originale. Afin d'étudier l'impact du choix de K sur la précision de la reconstitution de l'image, nous avons fait tourner plusieurs algorithmes knn avec un k allant de 1 à 10.

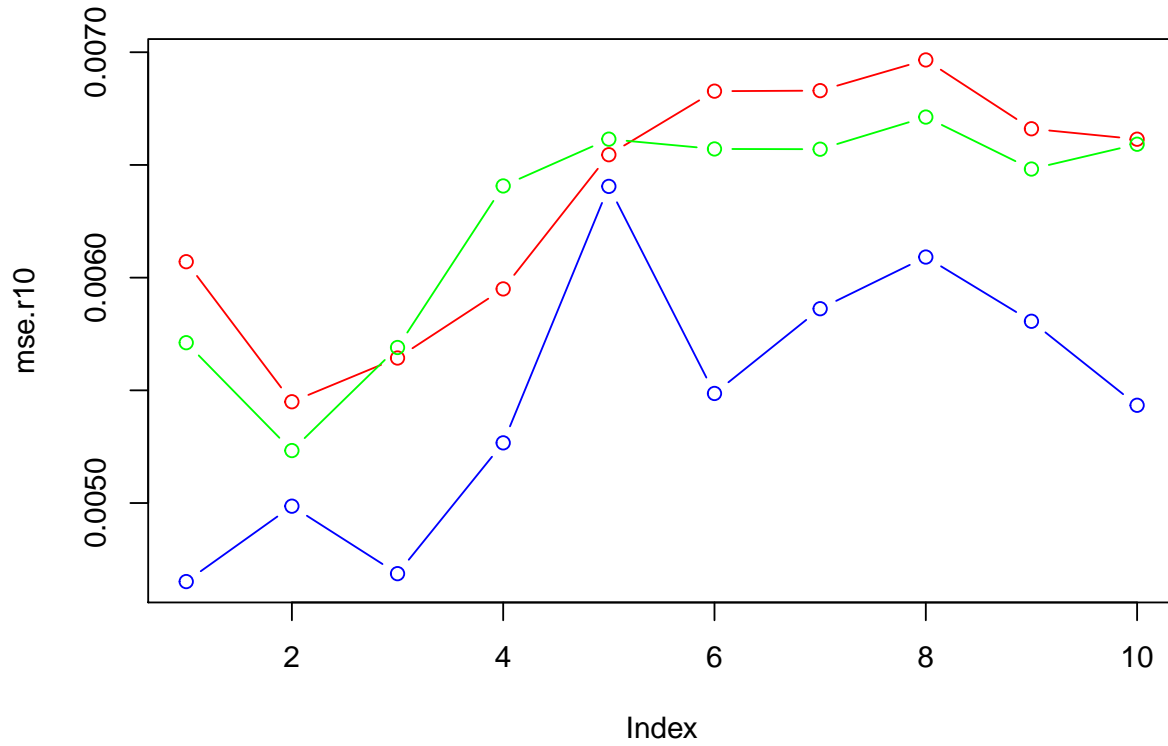


Figure 1: Evolution du MSE par couleur pour l'image avec 10% de bruit

On peut voir que l'évolution du MSE pour les couleurs rouge et verte est très proche, le K idéal semble être $k=2$. Alors que pour la couleur bleue il s'agirait plus de $k=1$ ou 3 .

#2. Etude de l'image avec 20% de bruit



A nouveau l'algorithme des knn avec $k=1$ semble efficace mais une étude de l'impact de K sur son efficacité est intéressante.

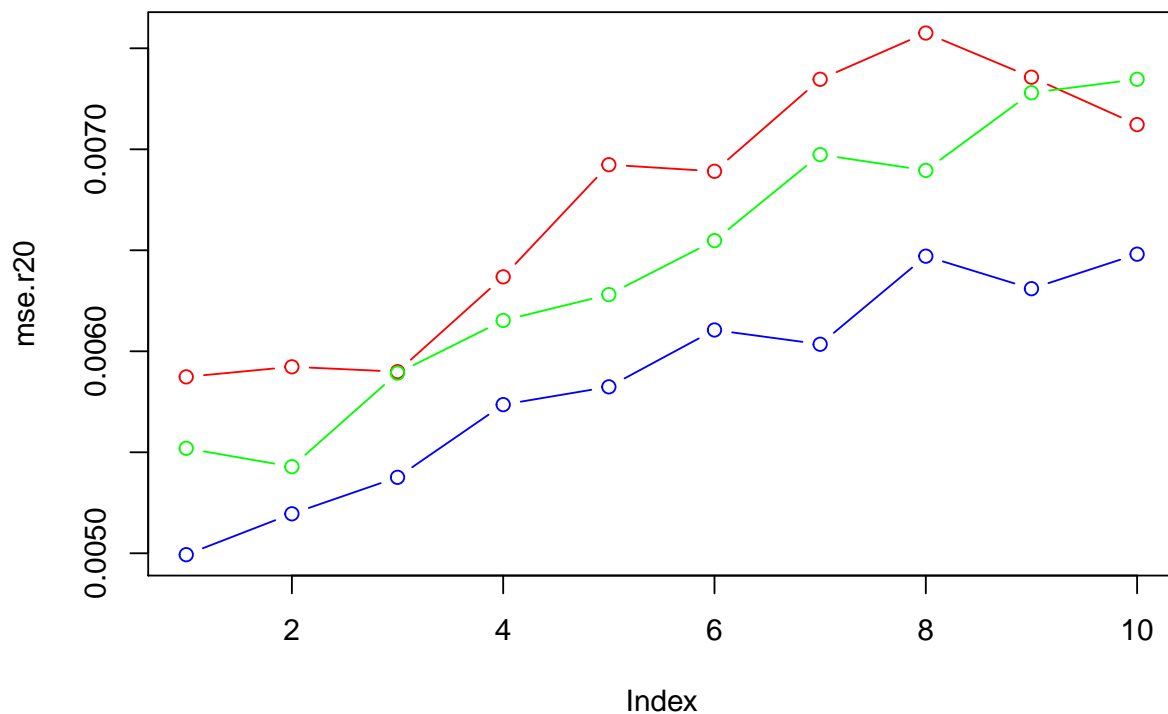


Figure 2: Evolution du MSE par couleur pour l'image avec 20% de bruit

A nouveau, seule la couleur bleue est quasi-linéaire, les couleurs rouge et verte sont les plus proches de la réalité lorsque $k = 2$ (ou 3 pour le rouge). Alors que pour la couleur bleue, plus K est élevé, plus l'erreur est grande.

#3. Etude de l'image avec 50% de bruit



Enfin sur cette dernière image, on voit clairement que l'image de droite n'est pas d'origine, mais lorsqu'on la compare avec l'image centrale, bruitée, l'algorithme paraît très efficace vu le résultat qu'il permet d'obtenir.

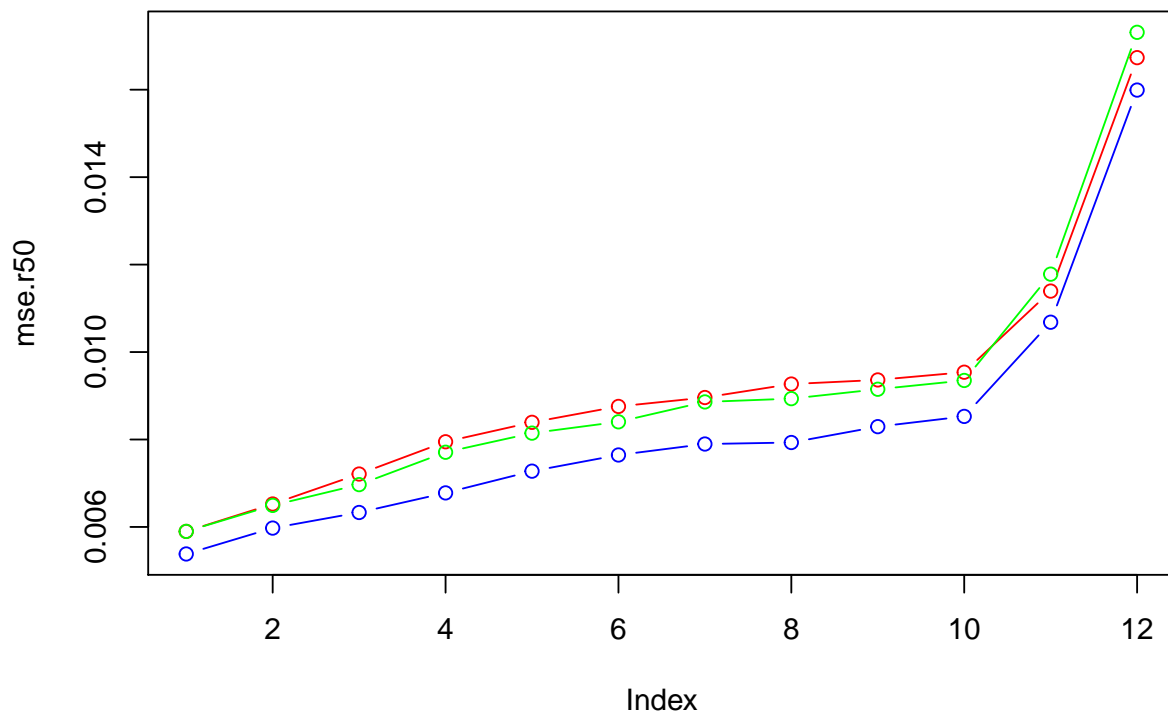


Figure 3: Evolution du MSE par couleur pour l'image avec 50% de bruit

Cette fois-ci l'évolution des MSE des trois couleurs se suivent beaucoup plus. L'erreur est minimale lorsque $k=1$, plus k augmente, plus l'erreur augmente aussi. Ces trois graphes auront donc permis de montrer l'évolution du MSE en fonction du nombre de K voisins choisis. On peut voir nettement que plus l'image a un bruit important, plus le nombre de K voisins choisis pour l'algorithme des knn doit être faible ou le MSE n'en sera que plus grand. On peut voir que l'évolution du MSE de l'image avec 20% de bruit n'a pas

la même évolution quasi-linéaire que l'image avec 50% de bruit, mais elle n'en est pas loin. Alors que pour l'image avec 10% de bruit, le choix d'utiliser qu'un seul voisin comme K n'est pas toujours le meilleur.

Nous avons donc choisis d'utiliser $K = 1$ pour les 3 images à reconstruire, étant donné qu'il s'agit du plus favorable aux 3 images réunis. Et obtenons les résultats suivants :

4. Impact du bruit sur l'efficacité de K

Enfin d'étudier l'impact du bruit sur le MSE en fonction de K fixé, nous avons choisis comme dit précédemment $K = 1$, étant donné qu'il s'agit du K entraînant le MSE minimal dans le plus de cas possible (2/3).

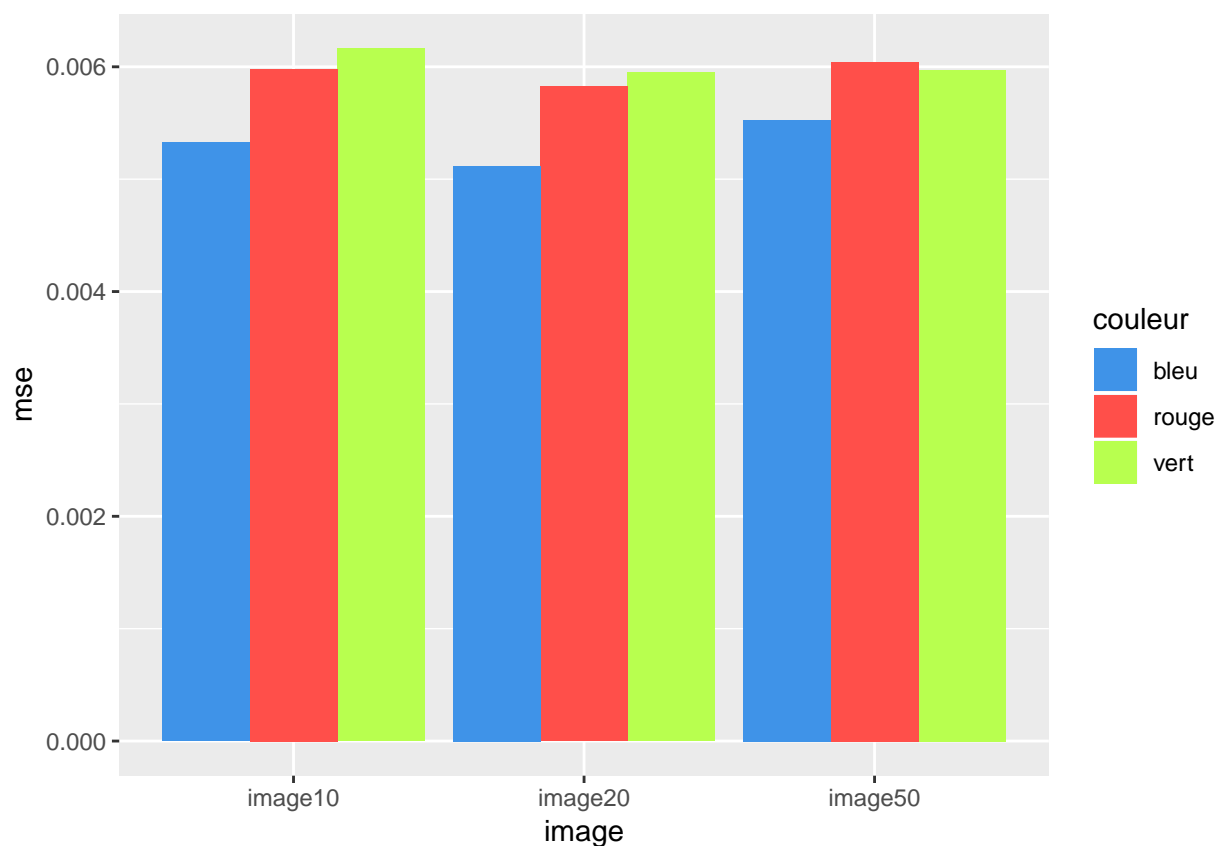


Figure 4: MSE par couleur et par image pour $K=1$

Le graphe précédent permet de voir que pour $K=1$ les MSE des trois images sont très proches malgré le bruit allant de 10 à 50%. De plus pour on aurait pu s'attendre à un MSE croissant, suivant le pourcentage de bruit. Pourtant l'image avec 10% de bruit, à un MSE concernant la couleur verte qui est supérieur au deux autres images. Cela montre que certaines couleur sont plus ou moins affectées pas le bruit. Le choix de $K=1$ n'étant pas optimal pour l'image à 10% justifie aussi cette différence.

```
## df.mse$image: image10
## [1] 0.0174686
## -----
## df.mse$image: image20
```

```
## [1] 0.0168907
## -----
## df.mse$image: image50
## [1] 0.0175416
```

En sommant les erreurs de chaque image on voit que l'image avec 50% d'erreur arrive quand même en tête, toutefois l'image avec 20% est loin derrière.

Conclusion

L'étude a permis de voir l'efficacité de la méthode des k plus proches voisins pour la restauration d'image, à l'oeil nu, lorsqu'il manque peu de pixel (environ 10%), la différence entre l'originale et la restaurée est faible. Au-delà on remarque la différence mais le résultat reste très acceptable. On peut voir par contre que le résultat à l'oeil nu est très loin de la réalité, en moyenne l'image à 10% est moins bien reconstituée que les images avec un pourcentage de bruit plus élevé. Cela montre peut être une des limites du MSE, qui serait de comparer des moyennes faites sur des effectifs différents. Le choix du nombre de voisins K est donc très important, même si on remarque que plus le nombre d'individus à prédire est important, plus K tend à être optimal en 1.

Annexe