

Problématique de la recherche d'images par le contenu Introduction

Sécurité d'image

Moteur de recherche

- Google possède 92,9 % du marché mondial dans ce domaine (contre 2,7 % pour Bing, et 2,2 % pour Yahoo!).
- **Requêtes Google**
- 30 000 milliards de pages sont indexées par Google.
- 20 milliards de sites sont visitées (crawlées) par Google, chaque jour.
- 3,3 milliards de requêtes sont effectuées chaque jour (100 milliards par mois).
- 15% des requêtes sont de nouvelles requêtes (500 millions par jour) !

Statistique YouTube

- Utilisateurs actifs mensuels (MAU) : 2 milliard
- Près de 5 milliards de vidéos sont visionnées sur Youtube chaque jour.
- Nombre de vues par jour sur mobile : 1 milliard (25% des vues)
- 6 milliards d'heures de vidéo vues par mois
 - Chaque minute, plus de 500h de vidéos sont uploadées sur Youtube !
- 100 ans de vidéos sont vérifiées chaque jour pour la gestion des droits

L'e-mail en chiffres

- **Nombre d'utilisateurs (source <http://www.arobase.org>)**

2,6 milliards en 2015.

Prévision pour fin 2017 : 2,9 milliards. Un tiers de la population mondiale utilisera alors l'e-mail.

- **Nombre de comptes e-mail**

4,4 milliards en 2015.

Prévision pour fin 2019 : 5,6 milliards

75 % des comptes sont personnels, 25 % professionnels

4

- **Nombre de mails envoyés chaque jour**

205 milliards en 2015 (hors [spam](#)).

Prévision pour 2019 : 246 milliards

le spam représente **entre 55 et 95 % du trafic total** de l'e-mail. La majeure partie (près de 90%) est filtrée en amont par les outils anti-spam des messageries.

	2015	2016	2017	2018	2019
Worldwide Email Accounts (M)	4,353	4,626	4,920	5,243	5,594
<i>%Growth</i>		6%	6%	7%	7%
Worldwide Email Users* (M)	2,586	2,672	2,760	2,849	2,943
<i>% Growth</i>		3%	3%	3%	3%
Average Accounts Per User	1.7	1.7	1.8	1.8	1.9

Table 1: Worldwide Email Accounts and User Forecast (M), 2015–2019

Chaque minute sur Internet

2019 *This Is What Happens In An Internet Minute*



2020 *This Is What Happens In An Internet Minute*



Introduction

- La recherche d'information dans les documents multimédias (en particulier l'image) est un problème crucial ...
- Elle a poussé les recherches dans le domaine de l'indexation et de la recherche d'image par le contenu
- Il existe deux type d'approche d'indexation :
 - Indexation sémantique (par mots clés)
 - Indexation descriptive (visuelle)

Indexation sémantique

- L'indexation manuelle attache au document des données de haut niveau relatifs à la signification du contenu de l'image.
- Il devient alors possible de comparer et de retrouver les images selon leur sémantique par l'intermédiaire de ces mots clés.
- Les requêtes associées sont en général des mots, désignant un objet, la date de prise, une action, le nom d'une personne ou d'un événement,...
- Une annotation textuelle sera toujours trop restrictive, même si elle prend en compte à la fois des informations syntaxiques et symboliques.
- **Avantage** : recherche indépendante du type de media

Indexation sémantique (3)



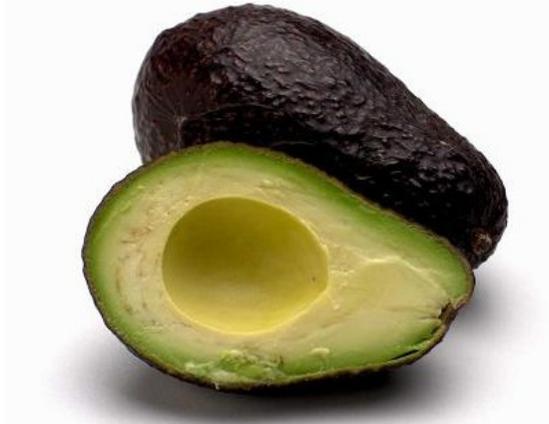
9

Indexation sémantique (3)

- **Inconvénient :**

- Difficile pour des bases d'images qui deviennent de plus en plus grandes
- Dépend de la langue utiliser
- Annotation manuelle
- subjective, elle dépend aux sentiments de chaque personne
- Différentes personnes utilisent un vocabulaire différent pour signifier la même chose “problème du vocabulaire” (ex: clair, lumineux).
- Le coût d'annotation manuel est très important

10

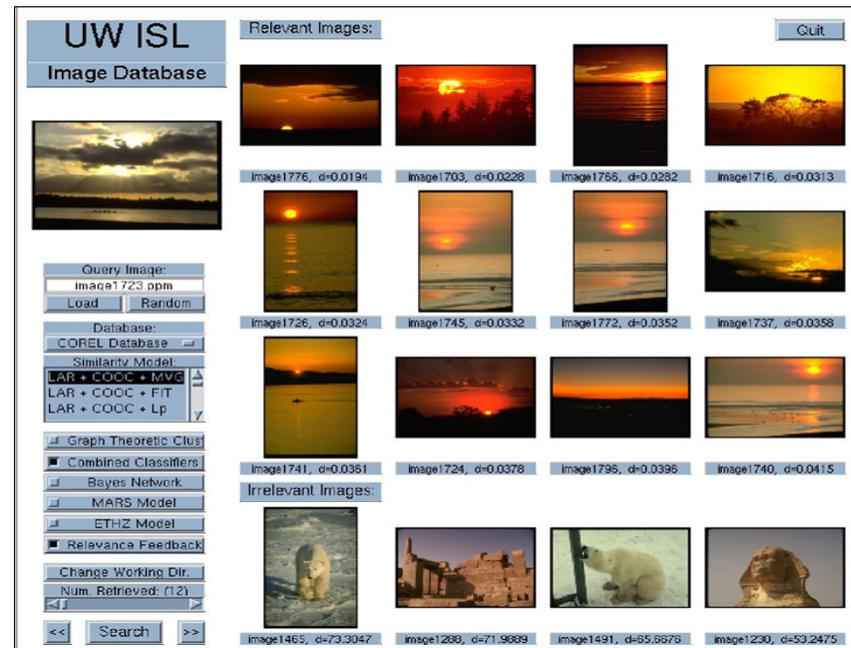
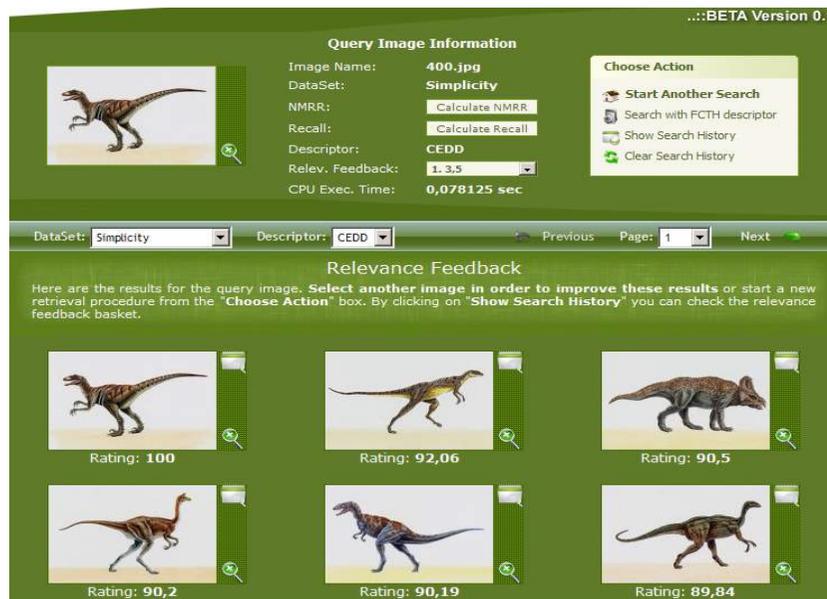


Avocat



Indexation d'images par le contenu

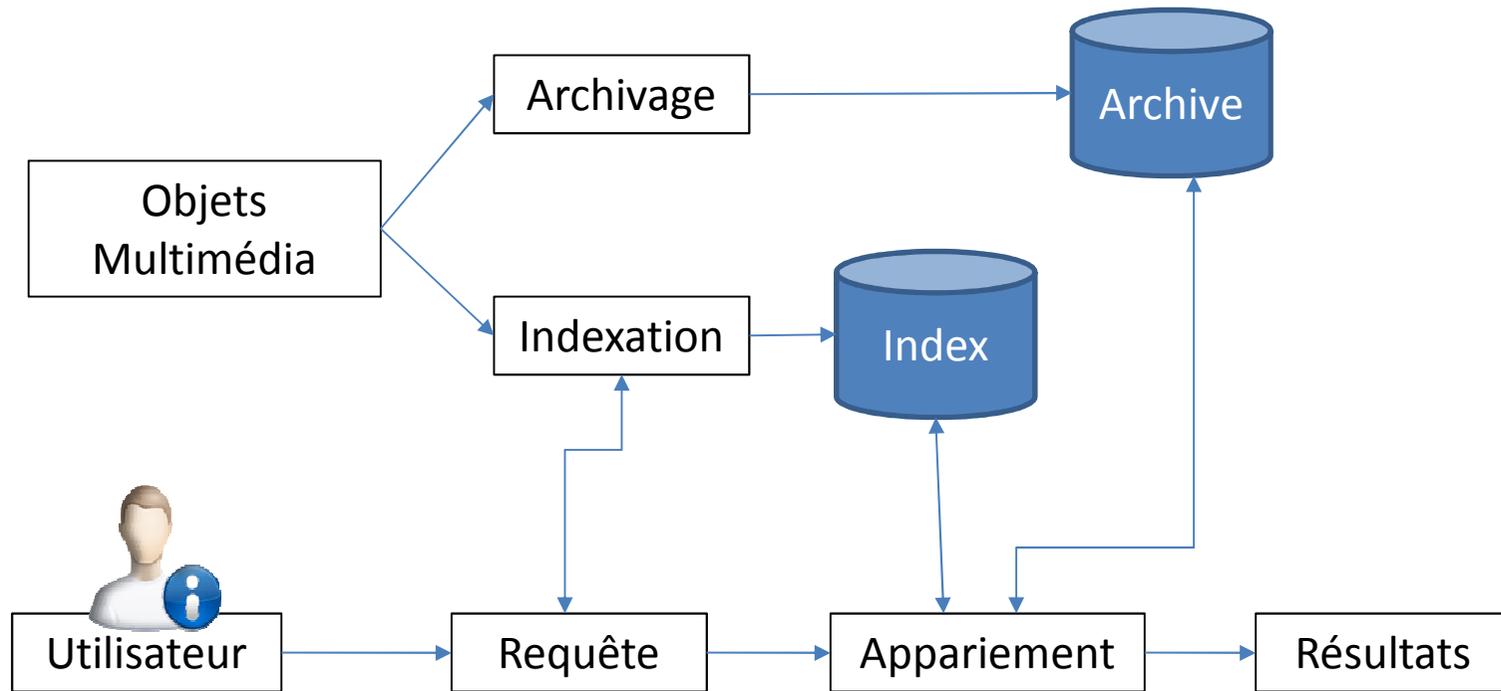
- L'indexation extrait automatiquement des caractéristiques de bas niveau, relatifs aux contenus de l'image spatiales, colorimétriques, géométriques ou statistiques, et les stockées dans un vecteur numérique appelé descripteur visuel.
- Les requêtes associées se font en général par l'exemple, partielle ou par modèle.



Différents domaines d'application

- Archivage, gestion des bases de données de contenu (personnelles ou professionnelles)
- Reconnaissance automatique du visage, d'empreinte humaine, d'objets, caractères...
- Détection de plagiat
- Protection du copyright
- Applications judiciaires : recherche de suspects
- Application médicales
- Application industrielles
- Commerce électronique
- Systèmes d'informations géographiques
- Moteur de recherche (ex: Internet)

Architecture générale

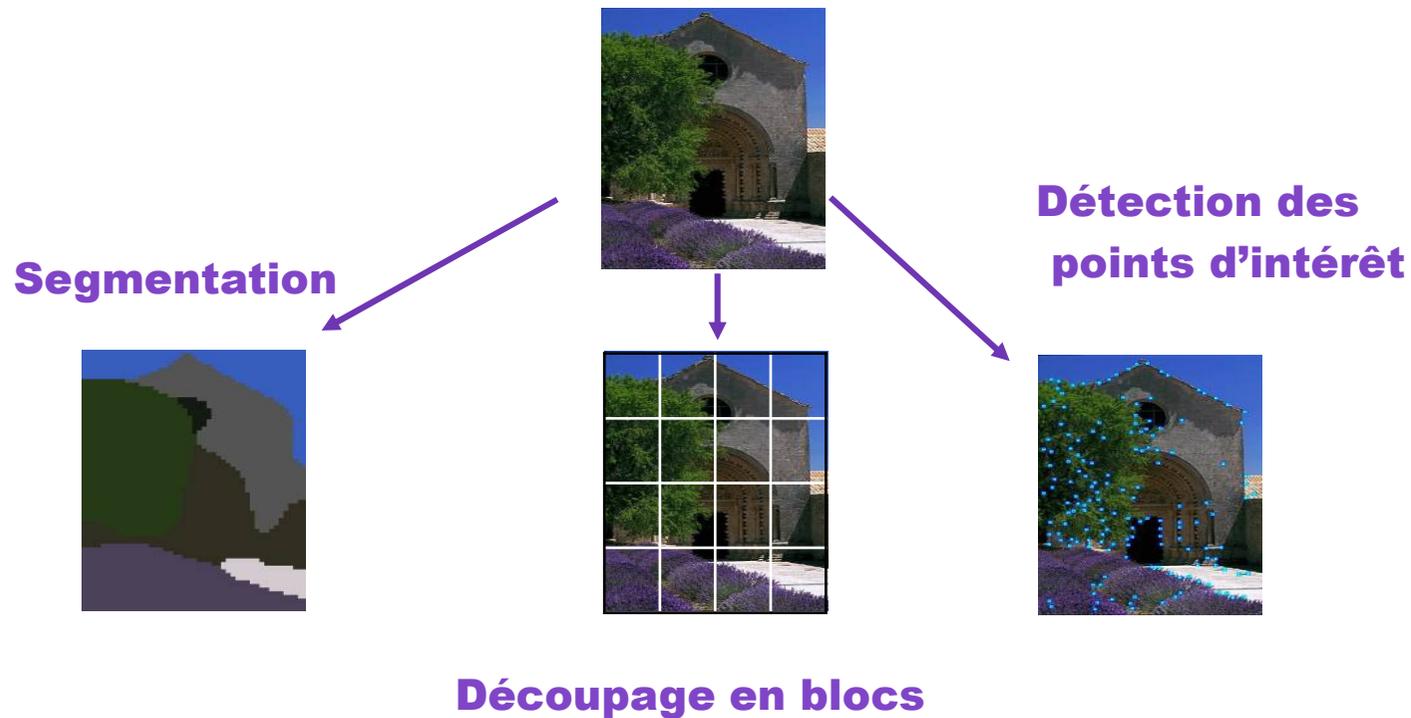


Indexation d'image

- Deux niveau d'indexation **locale** ou **globale**.
 - Indexation globale : l'approche usuelle pour caractériser une image consiste à calculer des caractéristiques globales sur toute l'image.
 - Indexation locale : consiste à calculer des caractéristiques sur des zones ou régions particulières de l'image.

Caractérisation locale

- L'importance de cette étape est d'identifier les zones particulières dans l'image pour le calcul des descripteurs.



Recherche d'images similaires

- Utile pour requête populaire
 - Beaucoup d'occurrence de la scène dans la base d'images.
 - Permet de retrouver toute les occurrences.



16

Recherche d'images ayant les mêmes propriétés

- Exemple : couleur similaires



Reconnaissance de visage : identification d'une personne

- Au sein d'une base de visages...



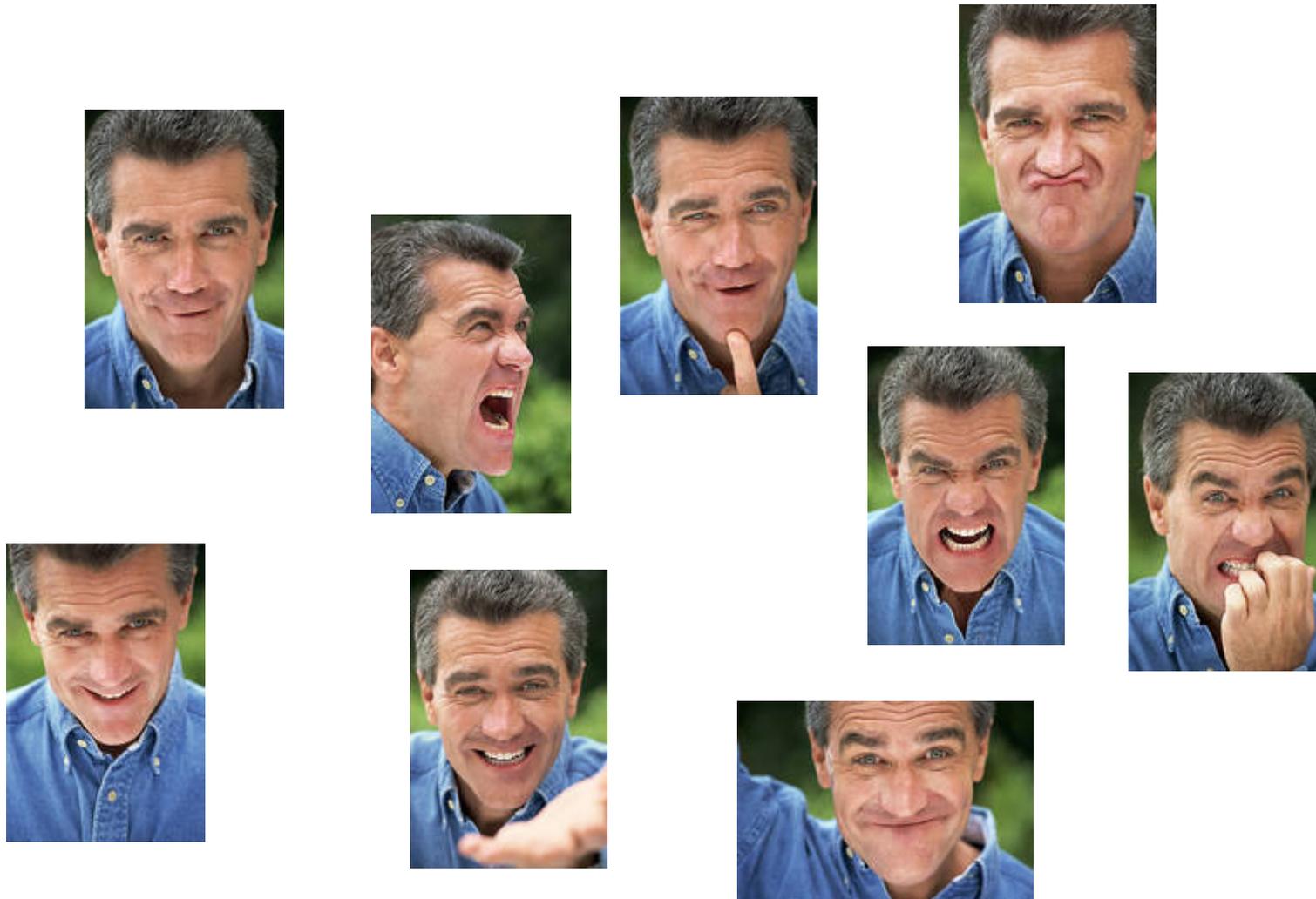
18

- ... ou à partir d'images ou de vidéos



Reconnaissance de visage : identification d'une personne

- Reconnaissance d'émotion ...



Classification d'images

- Reconnaître une classe d'objets (visages, voitures, etc.)
- Un objet en tenant compte de ses caractéristiques :



Classification d'images (2)

- Est plus complexe que rechercher des images de contenu similaire à une image requête
- Il requiert des informations sémantiques.



Problèmes rencontrés

- Contenu à reconnaître
 - Une image similaire
 - Une image contient un objet donné
 - Une instance d'une classe d'objets
- Tâches à effectuer
 - Segmentation (détection des bords de l'objet)
 - Détection (sous partie de l'image similaire, boîte englobante de l'objet)
 - Reconnaissance (image similaire, l'image qui contient l'objet ou la classe d'objet)

Problèmes rencontrés (2)

- La taille de la base d'images
- Transformations image : (changement d'échelle, rotation, luminosité, etc.)



- Image partielle / occultation



23

- Objet 3D



Descripteurs d'image

- Des attributs de différents types sont utilisés pour représenter le contenu de l'image. Les attributs sont classés en trois familles principales : la couleur, la texture et la forme.
- Les descripteurs de bas niveau les plus utilisés dans les systèmes actuels sont :
 - la couleur
 - la texture
 - la forme
 - Régions
 - Mouvement (pour la vidéo)

Descripteur couleur

- La couleur, avec son pouvoir discriminant, est un des descripteurs les plus employés en indexation d'images.
- Couleur moyenne, couleur dominante, distribution des couleurs
- **Problème** : choix de l'espace de représentation (RGB, HSV, YUV, etc.), distance entre couleurs
- Indice visuel insuffisant si utilisé seul, car peu discriminant :

25



Descripteur texture

- Une texture se caractérise par la répétition d'un motif de base dans différentes directions.
- On peut les séparer en deux classes: les textures structurées (macrotextures) et les textures aléatoires (microtextures).
- Deux types de méthodes
 - Statistiques : distribution spatiales des intensités de l'image
 - co-occurrences, auto-corrélation, ondelette
 - Géométriques : segmentation de motifs qui se répètent

26



Descripteur de forme

- La forme désigne l'aspect général d'un objet, son contour.
- Nombreux algorithmes d'extraction de formes : extraction des contours, les moments géométriques, etc.
- Difficultés :
 - l'étape d'extraction des contours est souvent instable
 - séparer la forme du fond est difficile (problème de segmentation)
 - leur calcul est relativement long et ils sont très sensibles au bruit.

Descripteur par région

- Se base sur
 - la répartition spatiale de zones « homogènes » + attributs aux régions (taille, couleur, etc.)
 - Les relations spatiales entre ces régions (ciel en haut/mer en bas)

28

- Sensible au bruit et aux occultations



La couleur

- La couleur, avec son pouvoir discriminant, est un des descripteurs les plus employés en indexation d'images.
- Différents espaces de représentation de la couleur ont été proposées (RGB, CIE XYZ, HSV, YUV,...).
- Avec la même méthode d'indexation, si on change l'espace de couleur, il peut donner des informations différentes de l'image.

Espace de couleur RGB

- L'espace RGB (Red-Green-Blue) (en français RVB) consiste à représenter chaque couleur par trois composantes le rouge (R), le vert (G) et le bleu (B).
- Est un espace additif, ses trois couleurs primaires se combinent d'une manière additive pour produire toutes les couleurs.
- chaque composante est codée sur 8 bits ce qui permet 256 niveaux d'intensité, de 0 à 255.
- Le principal inconvénient de cet espace de représentation vient du fait qu'il ne tient pas compte de la variation de la sensibilité de l'œil.

Espace de couleur YUV

- Le modèle **YUV** définit un espace colorimétrique en trois composantes. La première représente la luminance et les deux autres représentent la chrominance.
- Le signal YUV est créé depuis une source RGB (rouge, vert et bleu).
- YUV est utilisé dans les systèmes de diffusion télévisuelle PAL. Le système SÉCAM utilise quant à lui l'espace colorimétrique YDbDr, dérivé de YUV.
 - Y représente la luminance de la source.
 - U est obtenu en soustrayant Y du signal bleu.
 - V est obtenu en soustrayant Y du signal rouge.

Espace de couleur YUV (2)

- De RVB à YUV :

$$Y = 0,299 \cdot R + 0,587 \cdot G + 0,114 \cdot B$$

$$U = 0,492 \cdot (B - Y) = -0,14713 \cdot R - 0,28886 \cdot G + 0,436 \cdot B$$

$$V = 0,877 \cdot (R - Y) = 0,615 \cdot R - 0,51498 \cdot G - 0,10001 \cdot B$$

- De YUV à RVB :

$$R = Y + 1,13983 \cdot V$$

$$G = Y - 0,39465 \cdot U - 0,58060 \cdot V$$

$$B = Y + 2,03211 \cdot U$$

Où $(R, G, B, Y) \in [0 ; 1]^4$, $U \in [-0,436 ; 0,436]$

et $V \in [-0,615 ; 0,615]$

Espace de couleur HSV

- L'espace HSV ou **TSV** « teinte, saturation, valeur » est proche de la perception physiologique de la couleur par l'œil humain. Il consiste à décomposer la couleur selon :
 - La teinte (en anglais hue), correspondant à la perception de la couleur.
 - La saturation, décrivant la pureté de la couleur.
 - La luminance (en anglais value), indique la quantité de lumière de la couleur.
- la valeur varie entre 0 et 360, mais est parfois normalisée en 0–100 % ;

Espace de couleur HSV (2)

- Passage de RGB vers HSV

- max la plus grande valeur entre r , v et b ; et min la plus petite.

$$h \in [0, 360[\quad \text{et} \quad s, v, r, g, b \in [0, 1]$$

$$t = \begin{cases} 0, & \text{si } max = min \\ \left(60^\circ \times \frac{g - b}{max - min} + 360^\circ \right) \bmod 360^\circ, & \text{si } max = r \\ 60^\circ \times \frac{b - r}{max - min} + 120^\circ, & \text{si } max = g \\ 60^\circ \times \frac{r - g}{max - min} + 240^\circ, & \text{si } max = b \end{cases}$$

$$s = \begin{cases} 0, & \text{si } max = 0 \\ 1 - \frac{min}{max}, & \text{sinon} \end{cases}$$

$$v = max$$

Espace de couleur HSV (3)

- Passage de HSV vers RGB:
 - La conversion exploite les opérateurs "max" et "min".

$$h_i = \left\lfloor \frac{h}{60} \right\rfloor \bmod 6$$

$$f = \frac{h}{60} - h_i$$

$$l = v \times (1 - s)$$

$$m = v \times (1 - f \times s)$$

$$n = v \times (1 - (1 - f) \times s)$$

$$(r, g, b) = \begin{cases} (v, n, l), & \text{si } h_i = 0 \\ (m, v, l), & \text{si } h_i = 1 \\ (l, v, n), & \text{si } h_i = 2 \\ (l, m, v), & \text{si } h_i = 3 \\ (n, l, v), & \text{si } h_i = 4 \\ (v, l, m), & \text{si } h_i = 5 \end{cases}$$

Espace de couleur HSV (4)

- L'espace HSV est le plus utile pour la segmentation et la reconnaissance et il a été prouvé un espace très fort dans le système de recherche d'images.
- L'espace HSV n'est malheureusement pas uniforme, c'est-à-dire qu'une distance entre couleurs visuellement proches calculée dans cet espace peut être très grande.
- Un autre espace (Lab) a donc été créé pour résoudre cette difficulté.

36

Espace de couleur XYZ

- Matrice de passage RGB vers XYZ

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \frac{1}{b_{21}} \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} = \frac{1}{0.176,97} \begin{bmatrix} 0.490,00 & 0.310,00 & 0.200,00 \\ 0.176,97 & 0.812,40 & 0.010,630 \\ 0.000,0 & 0.010,000 & 0.990,00 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

37

- Matrice de passage XYZ vers RGB

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 0.418,47 & -0.158,66 & -0.082,835 \\ -0.091,169 & 0.252,43 & 0.015,708 \\ 0.000,920,90 & -0.002,549,8 & 0.178,60 \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$

Espace de couleur Lab

- Il a été spécialement étudié pour que les distances calculées entre couleurs correspondent aux différences perçues par l'œil humain.
- Il caractérise une couleur à l'aide d'un paramètre d'intensité correspondant à la luminance et de deux paramètres de chrominance qui décrivent la couleur .
- Le système Lab n'est pas directement relié au système RGB, on est obligé d'effectuer la transformation $RGB \Rightarrow XYZ$ avant de passer en CIE-Lab.

Espace de couleur Lab (2)

- Transformation $XYZ \Rightarrow Lab$:
- $L=116 \left(\frac{Y}{Y_n} \right)^{\frac{1}{3}} - 16$ pour $\frac{Y}{Y_n} > 0.008856$
- $L=903.3 - \frac{Y}{Y_n}$ pour $\frac{Y}{Y_n} \leq 0.008856$

- $a=500 \left(f\left(\frac{X}{X_n}\right) - f\left(\frac{Y}{Y_n}\right) \right)$

- $b=200 \left(f\left(\frac{Y}{Y_n}\right) - f\left(\frac{Z}{Z_n}\right) \right)$

- Où : pour $t > 0.008856$ $f(t) = \sqrt[3]{t}$
pour $t \leq 0.008856$ $f(t) = 7.7787t + \frac{16}{116}$

X_n , Y_n et Z_n correspondent au blanc décrit dans l'espace XYZ.

Espace de couleur Lab (3)

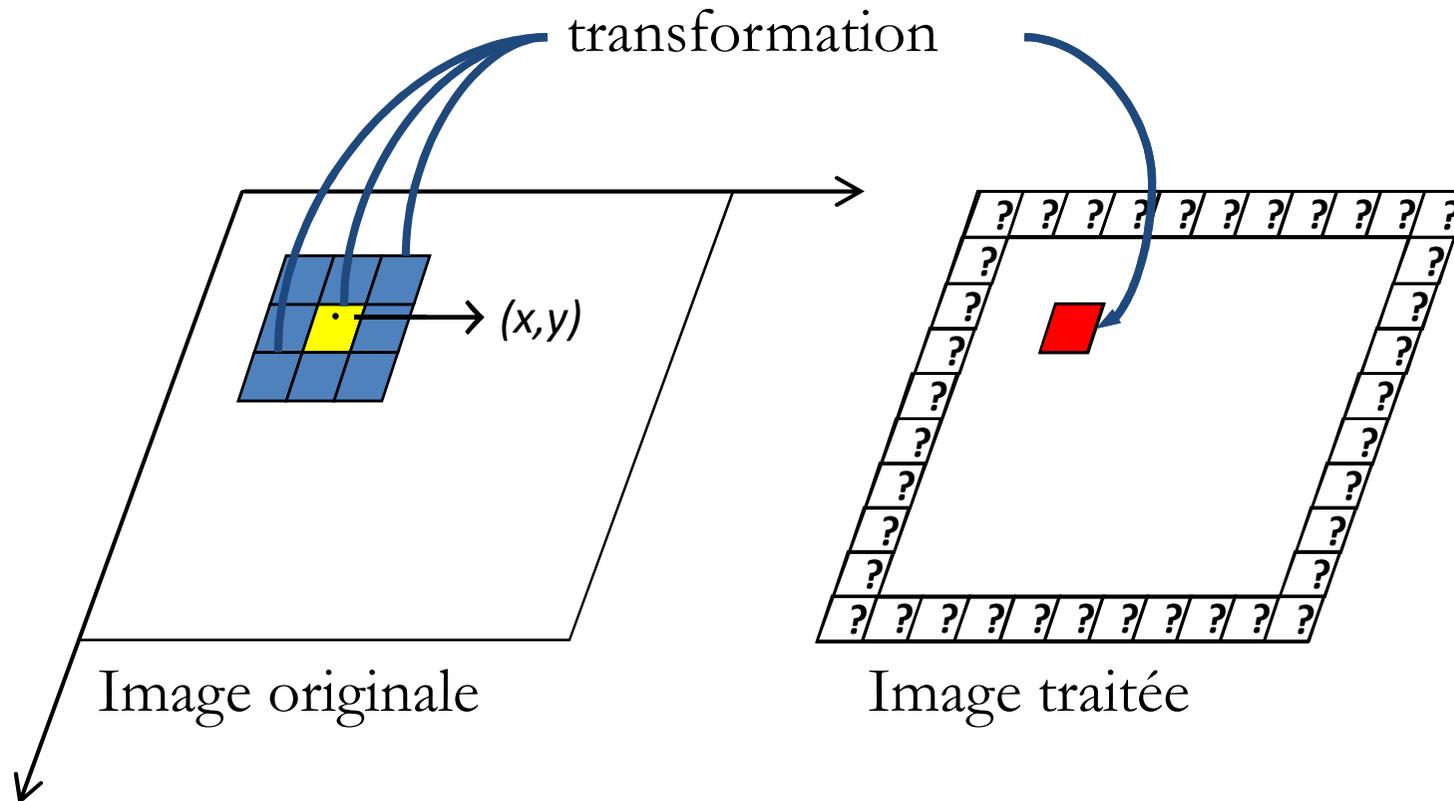
- Transformation Lab \Rightarrow XYZ :
- $Y = Y_n \cdot \frac{L}{903.3}$ pour $L \leq 8$
- $Y = Y_n \cdot \left(\frac{L+16}{116}\right)^3$ pour $L > 8$
- $X = Z_n \cdot f\left(\frac{a}{500} + \frac{L+16}{116}\right)$
- $Z = Z_n \cdot f\left(\frac{L+16}{116} - \frac{b}{200}\right)$
- Où : pour $t > 0.207$ $f(t) = t^3$
Pour $t \leq 0.207$ $f(t) = \frac{116 \cdot t - 16}{903.3}$

Discussion

- Le choix des espaces de représentation de la couleur adapté au traitement des images reste très ouvert et surtout très délicat.
- De nombreux chercheurs tentent de déterminer l'espace couleur le mieux adapté à leur problématique et leur travaux ont montré qu'il n'existe pas d'espaces couleur adéquat à toutes les problématiques de traitement d'images.

Filtrage spatiale

- Transformation basée sur le voisinage d'un point (x, y)



Convolution d'une image

- L'équation de la convolution d'une image par un filtre 2D est :

$$I'(i, j) = I(i, j) \otimes \text{filtre}(u, v)$$

$$I'(i, j) = \sum_u \sum_v I(i - u, j - v) \cdot \text{filtre}(u, v)$$

43

- En pratique, la convolution numérique d'une image se fera simplement par une sommation de multiplications.
- Note:
 - Généralement, le filtre est de dimension impaire et symétrique.
 - La plupart du temps, on divisera le résultat de la convolution par la somme des coefficients du masque.

Filtre passe-bas : filtre moyenne

Est obtenu simplement par convolution de l'image I, par le masque M suivant :

$$M = \frac{1}{9} \begin{vmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{vmatrix}$$

- Permet de lisser l'image.
- Réduit le bruit.
- Brouille ou rend floue l'image.
- Filtre dont tous les coefficients sont égaux.

44

Image normale

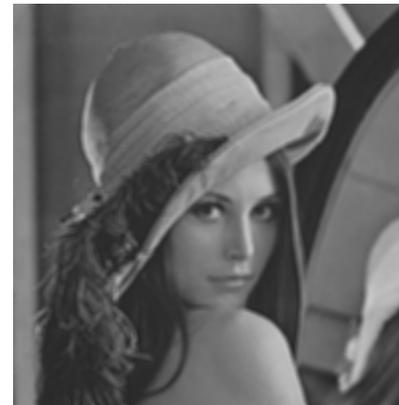
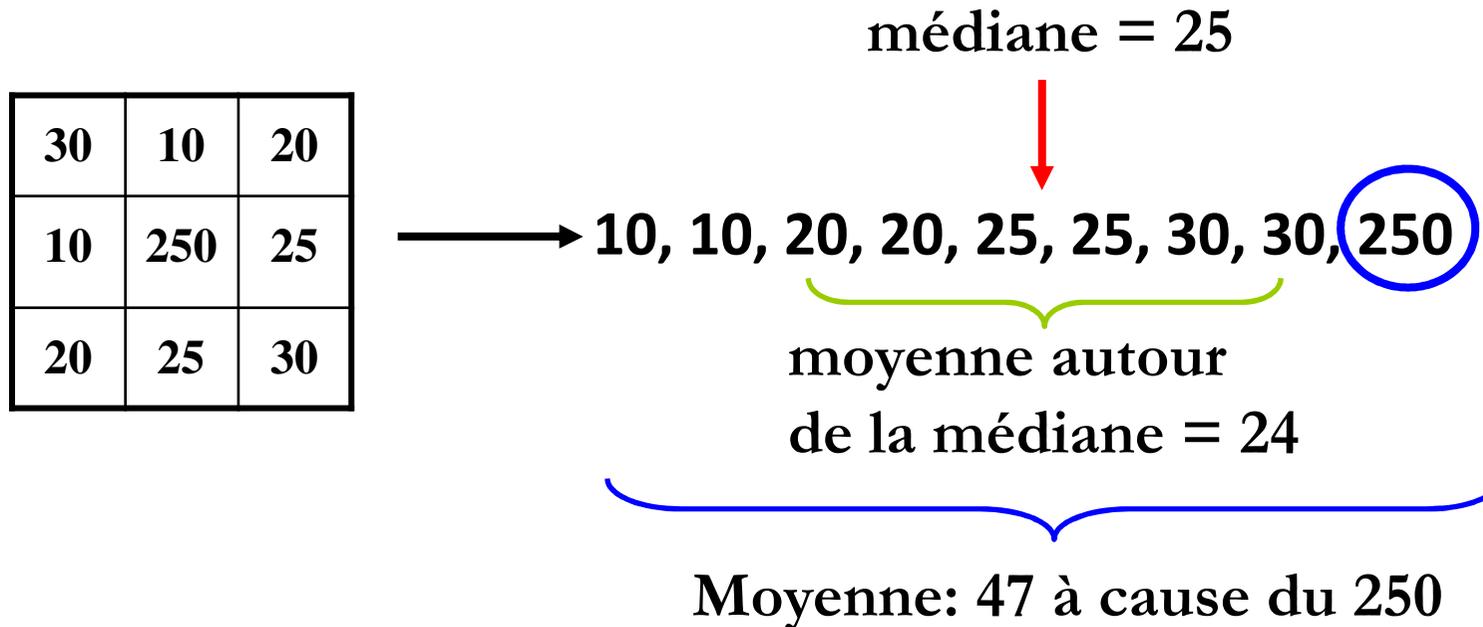


Image avec filtre moyen

Filtre médian



45

- Particulièrement utile pour un bruit de type Poivre & Sel (0 et 255)



Image originale



Bruit Poivre & Sel



Moyenne



Min



Max



Médian

Autres filtres passe-bas

Pyramidal:

$$\frac{1}{81} \times \begin{bmatrix} 1 & 2 & 3 & 2 & 1 \\ 2 & 4 & 6 & 4 & 2 \\ 3 & 6 & 9 & 6 & 3 \\ 2 & 4 & 6 & 4 & 2 \\ 1 & 2 & 3 & 2 & 1 \end{bmatrix}$$

Conique:

$$\frac{1}{25} \times \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 2 & 2 & 2 & 0 \\ 1 & 2 & 5 & 2 & 1 \\ 0 & 2 & 2 & 2 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

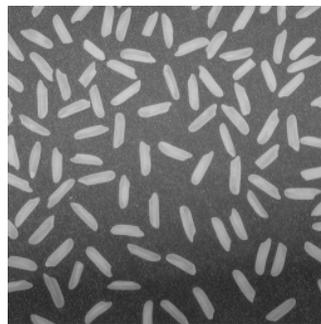
47

Gaussien:

$$\frac{1}{98} \times \begin{bmatrix} 1 & 2 & 3 & 2 & 1 \\ 2 & 6 & 8 & 6 & 2 \\ 3 & 8 & 10 & 8 & 3 \\ 2 & 6 & 8 & 6 & 2 \\ 1 & 2 & 3 & 2 & 1 \end{bmatrix}$$

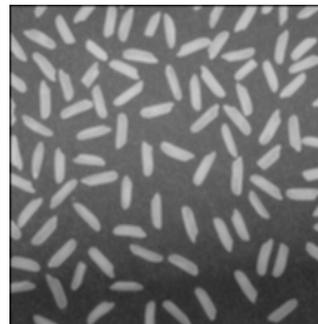
Filtre passe-haut

- Éliminer les fréquences basses
- Principe: Différence, Dérivation:
 - À partir d'un passe-bas
 - Gradient et variantes

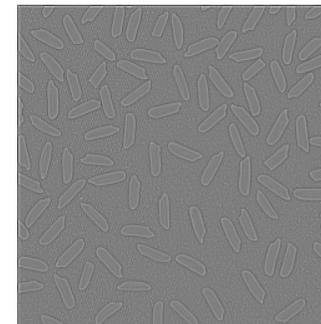


Image

0	0	0
0	1	0
0	0	0



Passe-bas

$$\frac{1}{9} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$


Passe-haut

$$\frac{1}{9} \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$$

$$\mathbf{I}_{PH} = \mathbf{I} - \mathbf{I}_{PB} \quad +128 \text{ pour affichage}$$

Le gradient

$$\nabla I = \begin{bmatrix} \frac{\partial I}{\partial x} \\ \frac{\partial I}{\partial y} \end{bmatrix} \quad |\nabla I| = \sqrt{\left(\frac{\partial I}{\partial x}\right)^2 + \left(\frac{\partial I}{\partial y}\right)^2} \approx \left|\frac{\partial I}{\partial x}\right| + \left|\frac{\partial I}{\partial y}\right| \quad \theta = \tan^{-1} \frac{\left(\frac{\partial I}{\partial y}\right)}{\left(\frac{\partial I}{\partial x}\right)}$$

$$\frac{\partial f}{\partial x} = \lim_{h \rightarrow 0} \frac{f(x+h, y) - f(x-h, y)}{2h} \Rightarrow \frac{\partial I}{\partial x} \approx \frac{I(x+1, y) - I(x-1, y)}{2}$$

$$\text{ou } \frac{\partial f}{\partial x} = \lim_{h \rightarrow 0} \frac{f(x, y) - f(x-h, y)}{h} \Rightarrow \frac{\partial I}{\partial x} \approx I(x, y) - I(x-1, y)$$

49

Le gradient (2)

Image

I_1	I_2	I_3
I_4	I_5	I_6
I_7	I_8	I_9

où I_x : valeur du niveau de gris

Filtre

0	-1	0
0	0	0
0	1	0

0	0	0
-1	0	1
0	0	0

Calcul

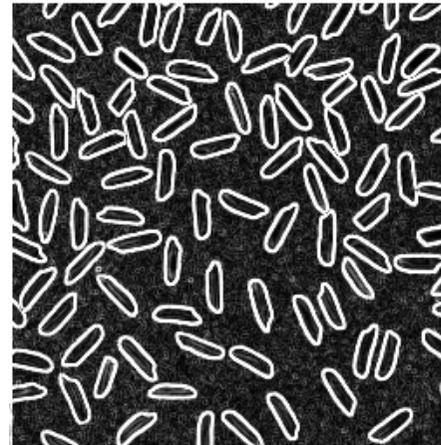
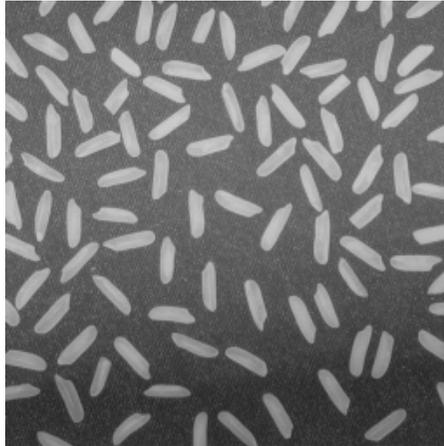
$$\frac{\partial f}{\partial y} = [I_8 - I_2]$$

$$\frac{\partial f}{\partial x} = [I_6 - I_4]$$

50

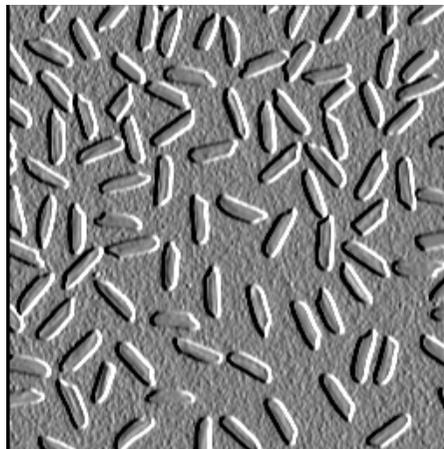
$$|\nabla I| \approx |I_x| + |I_y| \approx \left| \frac{\partial I}{\partial x} \right| + \left| \frac{\partial I}{\partial y} \right|$$

Exemple (sobel)



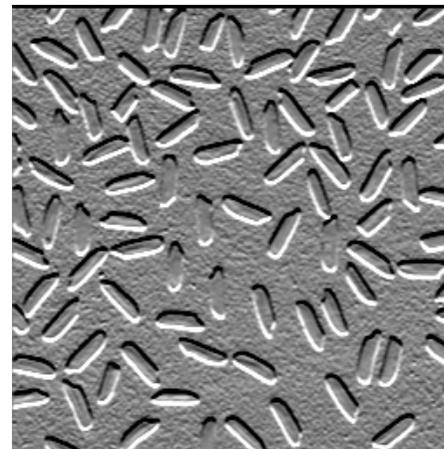
$$\nabla f = \left| \frac{\partial f}{\partial x} \right| + \left| \frac{\partial f}{\partial y} \right|$$

51



$$\frac{\partial f}{\partial x}$$

-1	0	1
-2	0	2
-1	0	1



$$\frac{\partial f}{\partial y}$$

-1	-2	-1
0	0	0
1	2	1

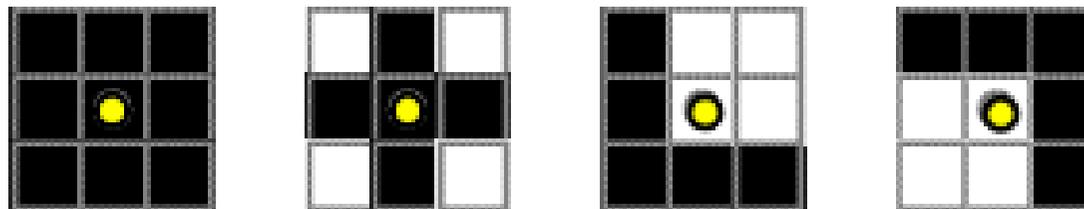
+128 pour affichage

Les opérations morphologiques

- Les algorithmes de morphologie sont appliqués sur des images binaire pour mettre en évidence leurs propriétés particulières.
- Ces algorithmes sont basés sur l'utilisation d'éléments structurants (notés B).

52

- **Élément structurant:**
- Un élément structurant est un masque binaire (constitué de pixels blancs et noirs) muni d'un point d'ancrage.



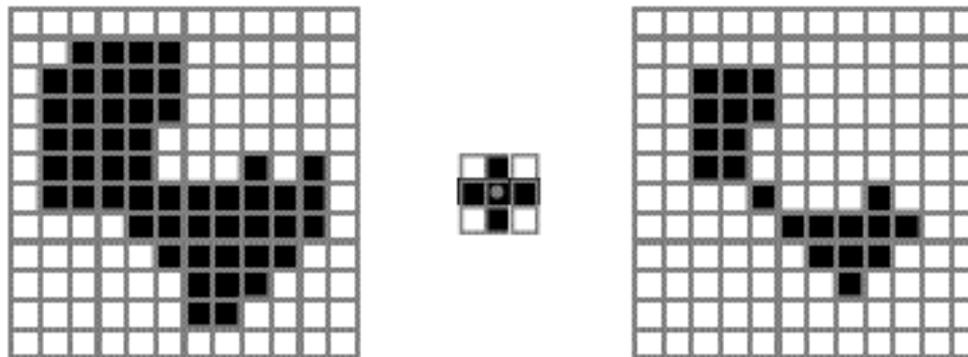
Érosion

- Pour chaque position de B sur l'image X, si tous les pixels de B font partie de X, alors l'origine de B appartient à l'image générée.
- Notation :

$$\text{Ero}_B(X) = \{(x,y) \mid B(x,y) \subset X\}$$

53

Exemple :



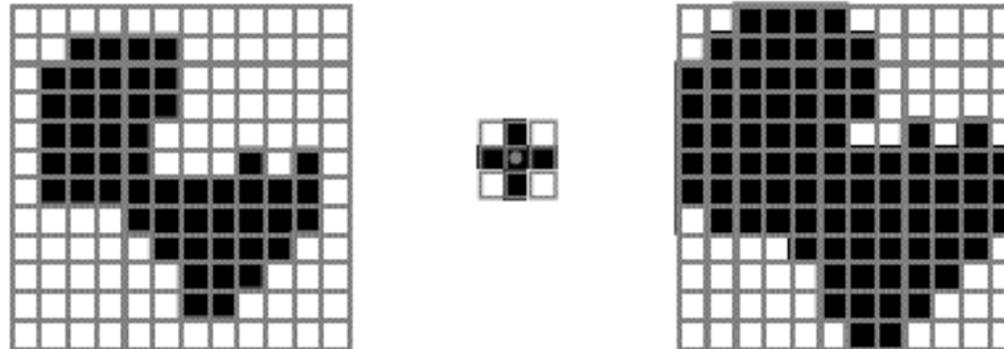
Dilatation

- Pour chaque position de B sur l'image X, si un, au moins, des pixels de B fait partie de X, alors l'origine de B appartient à l'image générée.
- Notation :

$$\text{Dil}_B(X) = \{(x,y) \mid B(x,y) \cap X \neq \emptyset\}$$

54

Exemple :



Ouverture

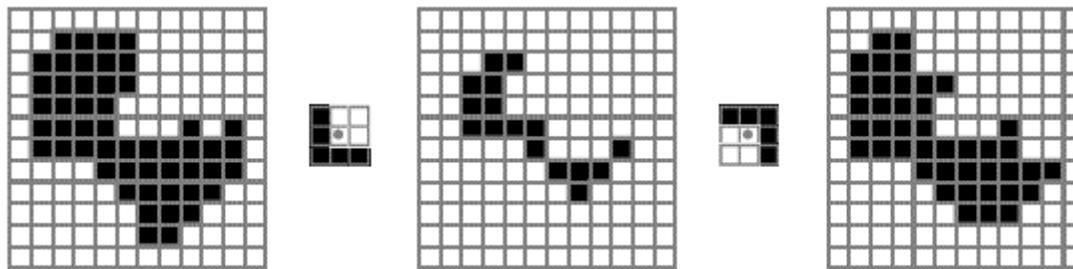
- L'ouverture de X par B est une image binaire définie par :

$$\text{Ouv}_B(X) = \text{Dil}_B(\text{Ero}_B(X))$$

- où B^- représente l'élément structurant symétrique de B par rapport au point d'ancrage.

55

Exemple :



Fermeture

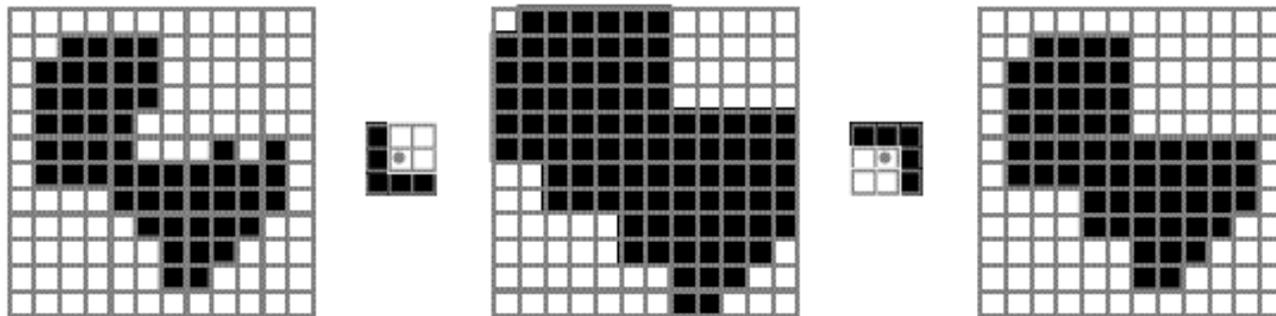
- La fermeture de X par B est une image binaire définie par :

$$\text{Fer}_B(X) = \text{Ero}_{B^-}(\text{Dil}_B(X))$$

- où B^- représente l'élément structurant symétrique de B par rapport au point d'ancrage.

56

Exemple :



Utilisation

- Les opérateurs morphologiques peuvent être utilisés dans plusieurs applications du traitement d'image:
 - l'élimination du bruit
 - l'extraction de contours
 - le remplissage de régions
 - amincissement et épaissement