

pour traiter de figures semblables, il faut renoncer à donner une réponse OUI ou NON et confectionner une *mesure de ressemblance*. Cette évaluation numérique va être donnée par ce qu'on appelle des distances : deux objets identiques seront à distance nulle ; deux objets très semblables seront à petite distance l'un de l'autre ; la distance entre deux objets très différents sera grande.

LA DISTANCE DE HAUSDORFF

Une première distance satisfaisante dans un assez grand nombre de cas est la distance de Hausdorff, nommée ainsi pour honorer Félix Hausdorff, mathématicien allemand mort en 1942. En 1914, Hausdorff a formulé les axiomes généraux de la topologie.

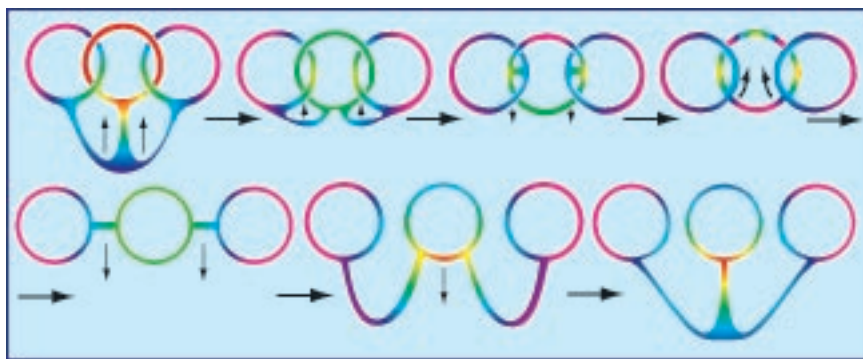
La distance de Hausdorff mesure la distance entre objets dessinés en noir et blanc sur une feuille de papier (des généralisations aux espaces à trois dimensions ou plus, ou prenant en compte les couleurs sont possibles). Pour définir cette distance entre figures, on utilise la notion usuelle de distance entre deux points.

Par définition, deux objets A et B dessinés sur une feuille de papier (c'est-à-dire deux ensembles de points A et B) sont à une distance de Hausdorff l'un de l'autre de moins de r unités, si chaque point de A est à moins de r unités d'au moins un point de B , et si, réciproquement, chaque point de B est distant de moins de r unités d'au moins un point de A .

Un procédé simple permet de visualiser ce qu'est la distance de Hausdorff entre deux ensembles A et B : on remplace chaque point de l'ensemble A par une petite tache ronde dont on fait grossir le rayon r jusqu'à ce que B soit recouvert par ce « A gonflé». On imagine aussi le processus réciproque (le recouvrement de A par un « B gonflé»). La distance de Hausdorff entre A et B est le plus petit r permettant simultanément les deux recouvrements.

Cette distance saisit correctement de nombreux aspects de l'idée intuitive de ressemblance. On considérera que deux dessins ne différant que par l'épaisseur des traits, ou par le fait que l'un est dessiné en pointillé, sont très proches l'un de l'autre : dès que l'on fait un peu gonfler l'un, il recouvre l'autre.

De même, la version brouillée d'un dessin A à travers un miroir dépoli sera peu distante du dessin A : de même encore, si, dans un dessin, vous remplacez les zones noires par des zones finement hachurées ou par des zones emplies de points aléatoires, alors la distance de la figure initiale à sa transformée sera petite.



3. Les transformations continues des surfaces réalisées en caoutchouc idéal permettent de passer d'une forme enlacée à une autre désenlacée, et «apparemment» sans rapport. Le semblable topologique apparaît trop tolérant.

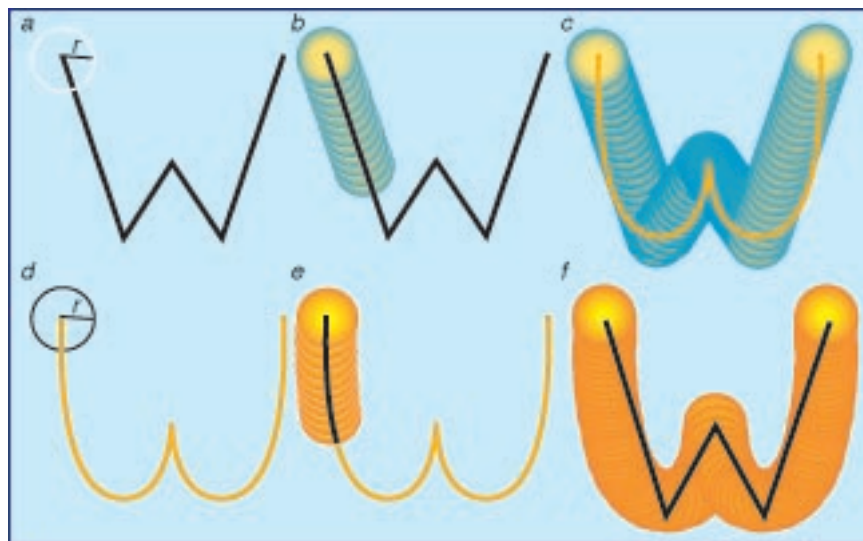
Un disque et une ellipse peu aplatie de rayons comparables seront proches pour la distance de Hausdorff. Un carré et un cercle, en revanche, seront plus éloignés. De façon générale, changer la texture ou déplacer légèrement les points d'un dessin en en préservant l'allure générale produit un dessin proche pour la distance de Hausdorff, alors que changer les formes crée des dessins éloignés.

Un cas intéressant est celui de l'approximation finie d'ensembles infinis. Un dessin comportant un nombre infini de points – par exemple, un disque plein – peut être approché de plus en plus finement par une série de figures ayant chacune un nombre fini de points. Ici l'infini, même non dénombrable, se laisse approcher facilement par du fini.

Cette distance est d'ailleurs implicitement utilisée par tous ceux qui dessinent des fractales : jamais personne ne dessine vraiment une fractale,

celle-ci étant, par définition, infiniment découpée, ce qu'aucun écran, aucun crayon, aucune imprimante ne réussit jamais à reproduire (notons aussi que personne ne dessine non plus de vrais cercles ou de vraies lignes). En revanche, les représentations qu'on donne des fractales sont bien des objets proches au sens de la distance de Hausdorff : comme Monsieur Jourdain faisait de la prose sans le savoir, nous utilisons la notion de ressemblance à la manière de Hausdorff sans jamais en avoir entendu parler.

Puisque deux dessins proches au sens intuitif sont mesurés proches au sens de la distance de Hausdorff, on comprend pourquoi cette distance est un outil essentiel de la morphologie mathématique, la discipline dont le but est la compréhension et l'analyse des images et des formes, et dont les applications concernent la vision artificielle, le traitement informatisé des images, la biologie et la géologie.



4. La distance de Hausdorff est le rayon r le plus petit tel qu'en remplaçant chaque point de A par un disque de rayon r on recouvre B , et inversement. C'est une vraie distance : $d(A,A) = 0$, $d(A,B) = d(B,A)$, $d(A,C) \leq d(A,B) + d(B,C)$ lorsque A , B et C sont des ensembles non vides et fermés (X est fermé si toute suite convergente de points de X a sa limite dans X).

INSUFFISANCE DE LA DISTANCE DE HAUSDORFF

Pourtant la distance de Hausdorff ne reconnaît pas certaines proximités entre formes que l'esprit humain identifie instinctivement. Ainsi la distance de Hausdorff entre un dessin et sa version en négatif (ou sa version où seul le contour des formes est gardé) est grande, alors que bien sûr nous, nous voyons très rapidement que le même objet est représenté. Pire, ajouter un unique point situé à l'extérieur d'un dessin donne un dessin très éloigné pour la distance de Hausdorff, alors que les deux objets nous apparaissent presque identiques. De même encore, la distance entre un objet et le même objet rapetissé d'un facteur constant est grande alors que nous percevons presque instantanément leur structure commune.

De même encore, la distance de Hausdorff ne permet pas de reconnaître la similitude entre des images d'un objet tridimensionnel (un visage, par exemple) vu sous divers angles ou des éclairages différents, alors que notre cerveau perçoit très vite l'analogie des formes.

Un défaut mineur, mais à signaler aussi de la distance de Hausdorff est qu'elle ne donne rien de bon avec des objets infinis en taille. Deux droites passant par un même point, ou bien sont confondues, et alors distantes de 0 (ce qui est satisfaisant), ou bien sont distantes de $+\infty$ même si leur angle est très faible (ce qui n'est pas acceptable).

Les déformations «caoutchouteuses» ne sont pas bien mesurées non plus : les montres molles de Dali ne sont pas proches des vraies montres pour la distance de Hausdorff, alors que, pour nous, elles le sont.

Remarquons qu'une combinaison de la notion de distance de Hausdorff avec la notion de similitude géométrique arrange certains de ces défauts : pour mesurer la distance entre A et B avec cette distance (appelée Hausdorff-bis), on envisage toutes les figures obtenues par application d'une similitude géométrique à B et l'on regarde celle qui, au sens de la distance de Hausdorff, est la plus proche de A ; c'est elle alors qui détermine la distance entre A et B pour Hausdorff-bis. Cette variante corrige les défauts concernant les déformations d'échelle et les déplacements, mais ne change rien aux insuffisances de la distance de Hausdorff concernant le passage au négatif, l'ajout d'un point à l'extérieur, les photos d'un même visage ou les déformations caoutchouteuses.

La dernière notion de ressemblance correspond à un progrès important sur le plan théorique, et en même temps

ouvre la porte à certaines applications pratiques. Cette notion, que nous appellerons distance informationnelle, est le résultat du travail commun de cinq chercheurs provenant de disciplines différentes : les physiciens Wojciech Zurek, du Santa Fe Institute, et Charles Bennett, du Centre de recherches IBM à New York ; le mathématicien Peter Gacs, de l'Université de Boston, aux États-Unis, et les informaticiens Ming Li, de l'Université de Waterloo, au Canada, et Paul Vitanyi, de l'Université d'Amsterdam.

LA DISTANCE INFORMATIONNELLE

La distance informationnelle présuppose que tous les objets que l'on considère sont des ensembles finis de points pris dans un ensemble discret. Il faut par exemple imaginer qu'il s'agit des pixels (noirs ou blancs) d'une image. Pour la distance de Hausdorff, une telle hypothèse de finitude n'était pas nécessaire : pour affiner un concept, on est parfois obligé de renoncer à l'infini, ce qui n'est peut-être pas étonnant : selon certains philosophes des mathématiques, l'infini n'est qu'une illusion. Des généralisations de la distance informationnelle à plus de deux dimensions et prenant en compte les couleurs sont possibles, mais nous ne les examinons pas ici.

L'idée fondamentale est que, si deux objets sont semblables, on passe facilement de la description de l'un à la description de l'autre, et réciproquement. En revanche, plus le passage de l'une à l'autre est long à détailler, plus les objets doivent être considérés comme différents, c'est-à-dire éloignés.

La difficulté du passage d'une description de l'objet A à la description de l'objet B est mesurée par la longueur du plus court programme qui transforme la donnée des points de A (supposés fournis pas énumération) à la donnée des points de B . Et donc la distance informationnelle est la somme des longueurs du plus court programme permettant de transformer A en B , et du plus court programme permettant de transformer B en A . On montre que cette distance varie peu quand on change le langage de programmation utilisé pour mesurer la taille des programmes.

L'idée d'utiliser une notion de plus court programme provient de la théorie algorithmique de l'information (appelée aussi théorie de la complexité de Kolmogorov), où l'on mesure la complexité d'un objet A par la taille du plus court programme qui produit A . La théorie de la distance informationnelle utilise d'ailleurs

de nombreux résultats de la théorie algorithmique de l'information.

Pour illustrer la notion de distance informationnelle, considérons le dessin de la tour Eiffel et le même dessin en négatif (nous avons dit que la distance entre ces deux objets pour la distance de Hausdorff est grande). Ces deux objets sont-ils proches pour la distance informationnelle ?

Oui, car on écrit facilement un programme de «passage au négatif» et que ce programme est court : le plus court programme transformant le premier dessin en le second (qui est le même ici que celui transformant le second en le premier) sera encore plus court (notons qu'il n'est pas indispensable de connaître précisément le plus court programme), et donc les deux dessins seront proches au sens de la distance informationnelle.

Ajouter un point à l'objet A donne un objet B que notre cerveau considère proche. En accord avec cette perception (et contrairement encore à la distance de Hausdorff, qui se trompait dans un tel cas), la distance informationnelle trouve que A et B sont très proches, car le programme le plus court ajoutant un point, et le programme le plus court enlevant un point (pour passer de B à A) sont tous les deux petits.

Plus remarquable et plus intéressant est l'exemple suivant : si l'on considère une image de la tour Eiffel prise selon un autre angle de vue, alors là encore on pourra par l'utilisation d'un programme assez court (qui contiendra des informations sur la position tridimensionnelle de la tour Eiffel) passer d'une image à l'autre. La parenté entre deux images provenant d'un même objet pris sous différents angles, ou même sous différents éclairages, est correctement détectée par la distance informationnelle.

Si l'on considère deux textes, même très longs, le second étant le résultat d'une remise en page du premier (changement des polices de caractères, des sauts des lignes, etc.), ils seront proches pour la distance informationnelle, car le passage de l'un à l'autre est décrit – et donc programmé –, de manière concise.

De nombreux autres exemples prouvent que cette mesure de distance fondée sur la taille des programmes satisfait l'essentiel de nos attentes, et que la grande majorité des défauts de la distance de Hausdorff disparaissent avec la distance informationnelle. L'utilisation pratique de cette distance semble cependant poser de graves problèmes.

Peut-on mesurer facilement la distance informationnelle entre objets ? Malheureusement, non. On majore sans mal