L'œil, la vision et la perception des images

Polycopié du cours 4IM06

version 3.50 : 28 avril 2025

Henri Maître
Département IDS, LTCI - Télécom Paris
henri.maitre@telecom-paris.fr

Objectifs de ce cours

Ce cours a pour objectif de faire comprendre comment la vision humaine permet de nous faire accéder à une représentation de l'univers qui nous entoure. Il explique pourquoi la perception visuelle est véritablement une étape de notre activité d'appropriation mentale et, à ce titre, contribue de façon privilégiée à notre conscience, c'est à dire à la relation intériorisée que nous nous faisons du monde. Nous verrons qu'en cela la perception visuelle est autant "poussée par les données" du flot optique que "tirée par les objectifs" de notre activité cérébrale consciente ou inconsciente.

Notre objectif est de montrer que, dans l'intention de créer une vision artificielle, il est probablement souhaitable de s'écarter du modèle qui enchaîne les traitements d'image en série et construit, sans retour en arrière, une interprétation de plus en plus riche (le paradigme de D. Marr qui a guidé la recherche depuis 40 ans), mais qu'il faut au contraire injecter très tôt "du sens" pour guider les traitements plus mécaniques.

*Les résultats assez impressionnants obtenus ces dernières années par les techniques à base de réseaux de neurones dans les domaines du traitement d'image et de la reconnaissance des formes qui s'appuient sur de grandes bases de données indexées (donc des signaux d'entrée "chargés de sens") semblent confirmer le bien-fondé de ce changement de paradigme. En retour, l'intégration de la vision et du traitement d'images au domaine génériquement nommé "Intelligence Artificielle" montre qu'il apparaît impossible aujourd'hui de concevoir une césure entre perception et raisonnement. Plus encore, l'étude fine de la vision se révèle comme une porte d'entrée exceptionnelle vers la compréhension de la conscience, clef de voûte d'une IA forte, comme le révèlent les travaux les plus récents des sciences de l'esprit (*mind-sciences*).*

Forts de cette conviction en une continuité intime entre perception et connaissance, nous pouvons décomposer le système visuel en ses fonctions élémentaires et examiner de quelles façons sont traités par la machine biologique, certains problèmes particulièrement délicats : détection des contours, suivi des mouvements, discrimination colorée, perception tridimensionnelle, persistance chromatique ... Pour cela, nous suivrons le signal optique depuis son entrée dans la cornée, sa capture par la rétine, son transport chimique et électrique par les nerfs optiques jusqu'aux aires visuelles et sa disparition ensuite (à notre connaissance aujourd'hui) vers les aires supérieures du cerveau où il interagit avec notre mémoire, nos émotions, nos intentions et les aires spécialisées, de la lecture, de la parole ... pour contribuer à une représentation complexe dite « phénoménologique », composante impalpable de la conscience.

Nous verrons comment deux modèles, à chacun des deux bouts de la chaîne, celui de D. Marr d'une part et celui des "gestaltistes" d'autre part, ont essayé de rendre compte de certaines étapes de ce traitement, fournissant des résultats très puissants pour aider à la conception des machines.

Nous examinerons les singularités de la vision qui révèlent nombre d'expériences largement médiatisées : aberrations, trompe-l'œil, images paradoxales ou impossibles, stéréogrammes aléatoires et autostéréogrammes ... Ces singularités, sur lesquelles notre système perceptif s'appuie en permanence au risque d'être très souvent trompé, éclairent les écarts qui existent aujourd'hui entre vision humaine et vision des machines.

Enfin, nous examinerons également la subtilité de la vision des couleurs et les espacex chromatiques qui ont été construits pour décrire mathématiquement la couleur, ainsi que les limites de ces formalisations.

Table des matières

1 L'Œil, la Vision et la Perception des images **7**

1.1 La formation de l'image chez l'humain 8

1.1.1 L'œil comme système optique 8

1.1.2 La rétine 9

1.1.3 Les voies optiques 13

1.1.4 Les aires visuelles 15

1.1.5 ... et ensuite ? 15

1.2 Quelques propriétés de la vision 18

1.2.1 L'œil et la caméra 18

1.2.2 Perception passive versus perception active 19

1.2.3 Une vision subjective 20

1.2.4 Des images sans vision 21

1.2.5 Un modèle algorithmique de vision : le modèle de David Marr 22

1.2.6 Une théorie de la perception : la *Gestalttheorie* ou psychologie de la forme . . . 23

1.3 Les illusions d'optique 24

1.4 La vision stéréoscopique 27

1.5 La couleur : modélisation et propriétés 28

1.5.1 L'espace RVB de la CIE 1931 29

1.5.2 La représentation XYZ 31

1.5.3 Quelques propriétés des espaces trichromatiques 33

1.5.4 Limites de l'espace RVB 34

1.5.5 Autres espaces chromatiques 35

1.5.6 L'espace Lab 36

1.5.7 Les limites de la colorimétrie 37

Bibliographie **39**

Chapitre 1

L'Œil, la Vision et la Perception des images

La vision est l'un des modes de perception les plus universels du monde animal. Elle tient également une très grande place pour l'espèce humaine. Des cinq sens majeurs, elle est certainement celui qui mobilise le plus notre système cognitif : une fraction importante (1/3 ?, 1/4 ?) de nos cellules cérébrales sont affectées au traitement des signaux visuels et l'œil apparaît, dès le stade fœtal, comme un organe singulier, extension de nos cellules cérébrales qui seront, par lui, exposées au monde extérieur. Dans la société, l'information visuelle échangée est extrêment importante et son rôle, comme il a beaucoup été rapporté, est croissant depuis un siècle [1].

Dans tout le règne animal, la vision exploite le rayonnement électromagnétique dans le domaine des longueurs d'onde optiques, depuis l'utra-violet proche (350 nm) jusqu'au proche infra-rouge (1 350 nm) [2] utilisant de façon très judicieuse le maximum d'énergie émise par le soleil (pratiquement un corps noir à 5 500 K) et une fenêtre de transparence de l'atmosphère La vision permet d'accéder à une représentation tridimensionnelle de l'univers, information nécessaire à la vie et à la survie des êtres animés. Au cours de centaines de millions d'années d'évolution, le monde animal a développé des organes de vision très différents selon les espèces. Les mammifères supérieurs ont tous à peu près les mêmes organes de vision et beaucoup d'autres espèces ont abouti des solutions proches (oiseaux, poissons, céphalopodes, etc.), c'est-à-dire une chambre optique formant une image sur une rétine sensible. D'autres animaux (mouche, caméléon, araignées, ...) ont développé des capteurs qui peuvent être très différents (mosaïques de récepteurs individuels sans optique, optiques par ondes guidées, ébauche d'images par sténopée ou par miroirs, etc.), mais pratiquement tous les animaux utilisent à peu près le même spectre de longueurs d'ondes et nous "partageons" donc un même monde visuel.

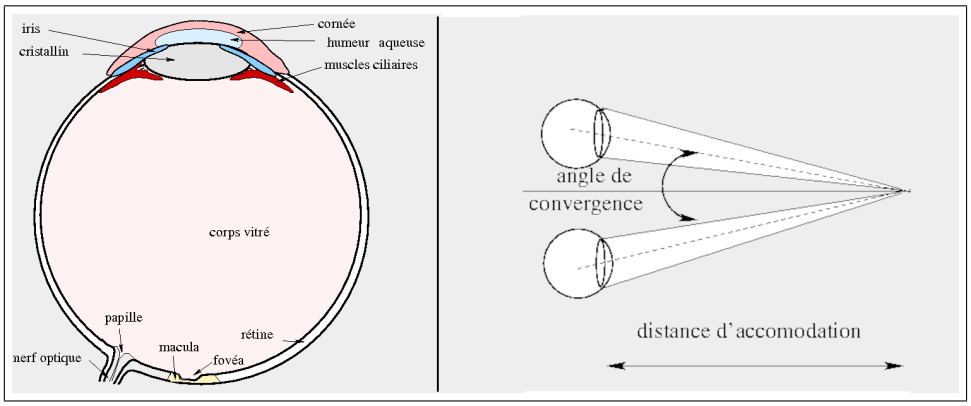

Figure 1.1 – *A gauche : les composants optiques de l'œil. La fovéa, partie centrale de la macula, est une zone sensible à haute résolution, la papille (ou tache aveugle) est la traversée du nerf optique. Le cristallin est la lentille déformable sous l'action des muscles ciliaires. A droite : distance d'accomodation et angle de convergence définissent le point de fixation du regard.*

1.1 La formation de l'image chez l'humain

1.1.1 L'œil comme système optique

Le capteur photo-sensible de l'œil est la rétine (figure 1.1). La formation de l'image sur la rétine est assurée par une lentille constituée du couple cristallin (déformable sous l'action des muscles du corps ciliaire ([Purves et al., 2015], chap. 11 ou [Le Grand, 1964, Wandell, 1995]) lors de l'opération dite d'**accomodation**) et cornée (solidaire de la déformation du cristallin). Le globe oculaire a un diamètre moyen de 15 mm. Lors de l'accomodation, l'image, inversée, se forme sur la rétine selon les lois de conjugaison des lentilles minces :

$$\frac{1}{p} + \frac{1}{p'} = \frac{1}{f} \tag{1.1}$$

La distance focale f de la lentille est liée aux courbures R_1 et R_2, à l'indice $n_c = 1,42$ de la lentille et à l'indice $n_v = 1,33$ du vitré (en contact avec le cristallin dans le globe oculaire) par la relation :

$$\frac{1}{f} = \left[(n_c - 1)\frac{1}{R_1} - (n_c - n_v)\frac{1}{R_2}\right] \tag{1.2}$$

L'œil moyen au repos a donc une puissance d'environ $\delta = 60$ dioptries soit une distance focale $f = 1,33$ cm.

C'est la variation de ces courbures sous l'action des muscles ciliaires qui provoque la modification de la longueur focale et donc de la distance de mise au point. Pour un œil sain, lors d'une contraction maximale des muscles ciliaires, la courbure est très faible, la mise au point de l'œil se fait alors à l'infini ou au *punctum remotum* (PR) en cas de myopie. Au maximum de courbure de la rétine, donc pour les objets proches, elle se fait au *punctum proximum* (PP). L'œil pourra voir net tout objet situé entre PP et PR, moyennant un effort d'accomodation. PP varie avec l'âge entre quelques centimètres à la naissance, une trentaine de centimètres à l'âge adulte et jusqu'à un mètre pour les seniors.

1. La vision est un sens très important et lorsqu'elle est altérée ou absente la vie individuelle et sociale peut se dégrader drastiquement. Il faut noter cependant que l'absence totale de vision dès la naissance apparaît moins handicapante que l'absence totale d'audition. Des palliatifs à l'absence de vision sont généralement mis en place lors de la croissance de l'enfant qui sont plus difficiles à trouver pour ce qui est de l'audition. L'insertion sociale des individus est souvent plus facile et meilleure dans les cas de cécité que dans les cas de surdité.

2. Pour l'homme comme pour beaucoup de grands mammifères, la plage perceptible est plus réduite et s'étend du violet (400 nm) au rouge profond (un peu moins de 800nm).

L'observation nette d'un objet à une distance donnée nécessite que l'**accomodation** soit faite pour chaque œil à cette distance, mais aussi que les deux yeux soient dirigés vers cet objet. Cela est obtenu par la **convergence** des deux axes visuels (figure 1.1 à droite). Convergence et accomodation sont des actions synchrones et réflexes (c'est-à-dire indépendante de notre volonté). Ces deux actions peuvent être rendues indépendantes par entraînement (par exemple par les opérateurs de stéréo-restitution).

Les images des deux yeux sont fusionnées selon un schéma complexe que nous verrons en Section 1.4, mais retenons dès à présent que les deux yeux donnent des images un peu différentes d'une même scène, différence fonction de la distance inter-oculaire. Cette distance (de 5 à 7 cm) crée une *base stéréoscopique* qui permet de séparer les divers objets de la scène en fonction de leur distance à la tête. L'information de *disparité* déduite de cette base stéréoscopique, est renforcée par les informations de convergence et d'accomodation pour nous permettre de reconstruire mentalement une scène tridimensionnelle. Les deux yeux n'ont pas tout à fait le même rôle dans notre perception du monde, l'un nous sert de « référence », c'est l'œil directeur (bien connu des chasseurs). C'est souvent l'œil droit pour les droitiers.

L'œil est un organe très mobile permettant, nous l'avons vu, d'assurer le pointage vers l'objet observé et déformable pour la mise au point sur cet objet à l'observateur. Mais l'œil est aussi l'objet de mouvements moins connus :

— des mouvements, volontaires ou réflexes mais conscients, les **saccades** qui permettent de maintenir l'objet d'intérêt sur la fovéa si l'objet est mobile, ou d'explorer une scène (voir la figure 1.2). Ces mouvements sont rapides (ils durent de 20 à 50 ms) et nombreux (de l'ordre de 3 par seconde) ;

— des mouvements totalement réflexes et inconscients, les **micro-saccades**, qui permettent de raffraîchir en permanence le signal d'image lors d'une opération de fixation. Leur amplitude est de quelques dizaines de minutes d'arc et se produit en quelques millisecondes.

L'iris est une membrane teintée, située en avant du cristallin, percée en son centre d'une ouverture circulaire de taille variable, la pupille, permettant d'adapter le flux de lumière arrivant sur la rétine en fonction de l'éclairage de l'objet. La cornée est une membrane transparente qui protège l'œil vers l'extérieur. Le globe oculaire est empli d'un liquide gélatineux, le corps vitré, qui donne sa rigidité à l'œil. La rétine tapisse toute la surface interne de la chambre oculaire, à l'exception d'une petite zone où débouche le nerf optique : la papille ou zone aveugle. Remarquons que la rétine ne forme pas un "plan-image" mais une "surface-image" sphérique qui donne une géométrie de l'image aux angles forts différente de celle que l'on considère généralement en optique avec des lentilles minces ou avec un appareil photo (la formule 1.2 n'est valable qu'au voisinage' de la fovéa).

L'œil a une résolution d'une minute d'arc (0,017 $^\circ$) environ, mais seulement sur une toute petite partie de sa surface sensible, sur la **fovéa** qui ne sous-tend qu'un angle solide de 1,5 $^\circ$ environ [3]. Cette zone à très haute résolution correspond donc grossièrement à une image circulaire de 100 pixels de diamètre. Le champ visuel global n'est pas circulaire, il est de 90 $^\circ$ du côté temporal, de 60 $^\circ$ du côté nasal, de 50 $^\circ$ et 80 $^\circ$ vers le haut et vers le bas. La zone à haute résolution n'est donc qu'une petite fenêtre à l'intérieur d'une image beaucoup plus vaste dont la résolution se dégrade rapidement hors de la fovéa. Néanmoins, la disponibilité d'un très vaste champ visuel, même s'il est à faible résolution, est d'une très grande importance pour l'analyse d'une scène (et pour la survie en traversant une rue !).

1.1.2 La rétine

Les rayons lumineux qui ont traversé le vitré sont absorbés par la rétine. La rétine est composée d'environ 5 millions de cônes, chargés de la vision diurne et colorée, et de 100 millions de bâtonnets, chargés de la vision nocturne, de la détection des lumières faibles et des changements rapides.

Les bâtonnets captent la lumière à l'aide d'un seul pigment : la rhodopsine dont le maximum de

3. Il ne faut cependant pas limiter la résolution extrême de la vision à 1 minute d'arc car de nombreuses expériences prouvent que l'utilisation des micro-saccades permet de voir des détails beaucoup plus fins sur des objets contrastés : par exemple un cheveu ou un fil de soie. Les performances du capteur sont surpassées grâce à une boucle active qui met en œuvre l'ensemble du système visuel dans une expérience dynamique.

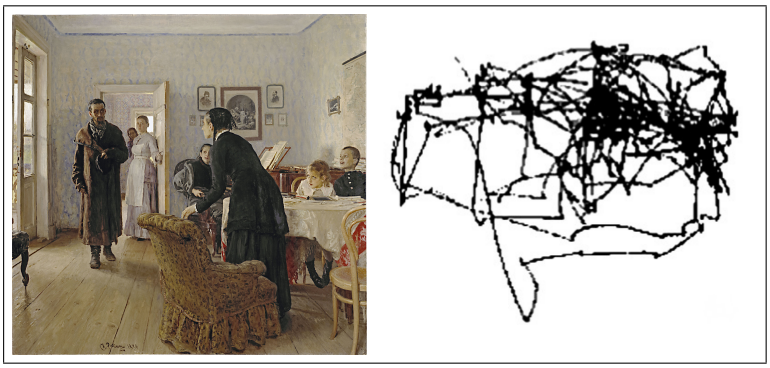

Figure 1.2 – *Chemin que suit l'axe du regard (d'après Yarbus et Kolers [Yarbus, 1967]) lorsqu'on observe une image (ici une peinture d'Ilya Répine). Ces déplacements du regard sont causés par les saccades tandis que les petits mouvements, autour des positions d'intérêt sont dûs aux micro-saccades.*

sensibilité est au voisinage de 500 nm (figure 1.3). Les bâtonnets sont répartis de façon très dense en périphérie de la rétine, puis leur densité diminue jusqu'à s'annuler au centre de la fovéa.

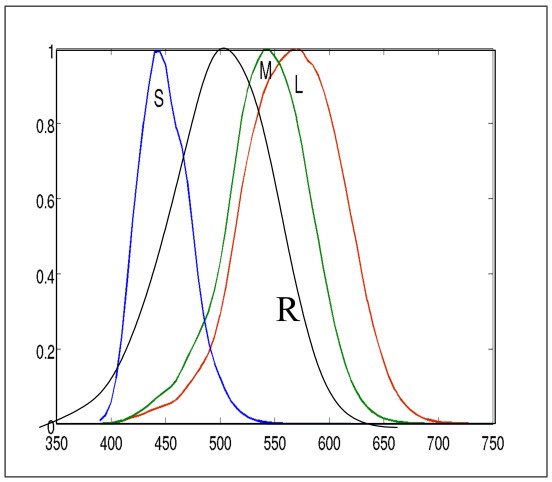

Figure 1.3 – *Courbes normalisées de la réponse des trois types de cônes, en fonction de la longueur d'onde, en nanomètres, pour un angle de vue de deux degrés (les propriétés de la vision se mesurent toujours par intégration sur une petite plage du champ visuel, or, la rétine n'étant pas homogène), ces propriétés varient selon la taille du champ observé. Toutes les mesures sont donc toujours spécifiques à un angle de vue donné). On note la grande proximité entre les réponses des cônes M et L. La courbe notée R est celle de la rhodopsine des bâtonnets.*

Les bâtonnets sont très sensibles. Ils réagissent à l'absorption d'un seul photon, mais lorsque l'éclairement croît, ils sont rapidement saturés. Les cônes sont beaucoup moins sensibles (environ 200 fois) mais disposent d'une dynamique beaucoup plus grande (Figure 1.9). Cependant la totalité de la dynamique n'est pas disponible instantanément. Les cônes fonctionnent "autour d'un point moyen" ; exposés à des éclairages plus forts ou plus faibles, ils font évoluer ce point moyen avec des temps d'adaptation relativement brefs (de l'ordre de la minute). Par contre les bâtonnets, lorsqu'ils ont été saturés (comme en plein jour) puis reviennent dans l'obscurité, ont besoin d'une vingtaine de minutes d'adaptation pour retrouver leur dynamique.

Il y a trois sortes de cônes différemment sensibles aux longueurs d'onde et responsables de la vision des couleurs [4]. Les maximums de leur sensibilité sont situés respectivement à 440 nm (dans le

4. Notons également que si l'homme a trois types de cônes (à l'exception des daltoniens (voir figure 1.4) qui n'ont que deux types de cônes, voire parfois un seul, de nombreux animaux n'en ont que deux comme le chat, ou un seul

bleu), 530 nm (dans le vert) et 560 nm (dans le jaune-vert) (figure 1.3). Ils sont dénommés de diverses façons :

— la dénomination anglo-saxonne utilise les qualificatifs de *long*, *middle* et *short* (LMS), ces termes qualifiant la position, en longueur d'onde, du maximum de leur sensibilité. Le langage courant s'accorde à retenir cette dénomination LMS que l'on traduit en français par *long*, *moyen* et *court*.
— la dénomination savante est cônes cyanolabes, chlorolabes et érytholabes,
— une dénomination fréquente utilise les termes RVB, conformes à ce que l'on utilise en imagerie électronique, quoiqu'ils en soient assez éloignés comme nous le verrons plus loin.

La vision des couleurs fera l'objet de la Section 1.5.

Les cônes L sont environ deux fois plus nombreux que les M et dix fois plus que les S [5].

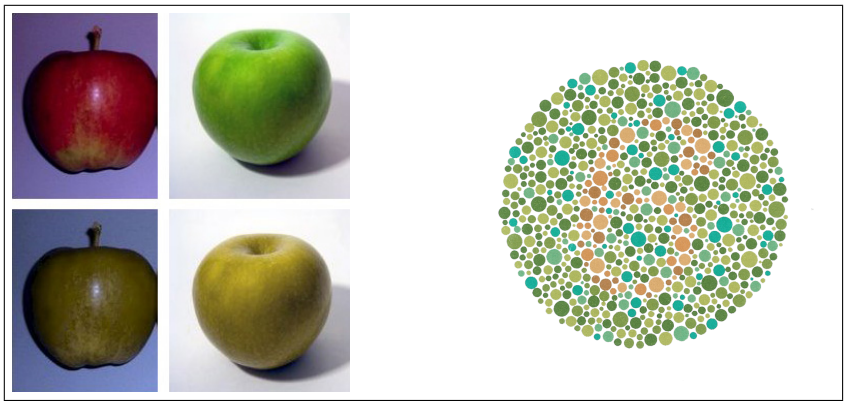

Figure 1.4 – *Anomalies de la vision dues au défaut de certains cônes : le daltonisme. L'absence de l'un ou de deux pigments parmi les trois (LMS) explique les six sortes différentes de daltonisme chez l'homme. Une septième anomalie, l'achromatopsie, correspond à une vision monochrome à l'aide des seuls bâtonnets. A gauche : en haut, deux pommes telles qu'elles sont vues en vision trichromatique. En bas, telles que vues en vision bichromatique deutéranope (absence des cônes M) (simulation : H. Brettel et al., 1994, Wikipedia). A droite une page de l'album de tests d'Ishihara permettant de détecter les anomalies de la vision colorée.*

Des trois signaux issus de ces capteurs naît la trivariance visuelle, reconnue par Grassman en 1853, bien avant que l'on identifie les cônes et les bâtonnets : « tout stimulus coloré peut être reproduit par un mélange additif de trois primaires convenablement choisies ». Nous examinerons plus en détail cette trivariance et ses conséquences à la section 1.5.

Les cônes sont répartis sur une grille assez régulière au fond de la rétine, et non à sa surface comme on pourrait le croire. La lumière doit donc traverser toute la rétine avant d'être absorbée par les pigments situés à l'extrémité des cellules réceptrices. Ce montage surprenant s'explique par le double rôle joué par l'épithélium pigmentaire dans lequel sont plongées les extrémités des cônes et des bâtonnets. L'épithélium partipe tout d'abord au renouvellement des pigments rétiniens après une exposition à la lumière. Il assure ensuite l'élimination des disques membraneux qui terminent les cônes et les bâtonnets et qui contiennent les photorécepteurs car ces disques s'épuisent très vite (en quelques jours). Ces disques se détachent alors ; l'épithélium pigmentaire est chargé de les évacuer et les faire disparaître vers l'extérieur de l'œil.

Les réponses spectrales de la figure 1.3 indiquent la probabilité qu'un photon de longueur d'onde donnée soit absorbé. Lorsqu'un photon est absorbé par un cône, quelle que soit sa longueur d'onde, il donne naissance à un potentiel électrique de membrane grâce à un pigment nommé *iodopsine*. Ce potentiel est transmis à une terminaison synaptique où un neuro-transmetteur est libéré. Le signal chimique est converti en signal électrique lors de l'ouverture de canaux ioniques dans la zone post-

comme le rat, tandis que le pigeon en a cinq.

5. L'existence d'un quatrième type de cône (sensible dans l'orange), présent dans une partie de la population, fait encore l'objet de spéculations.

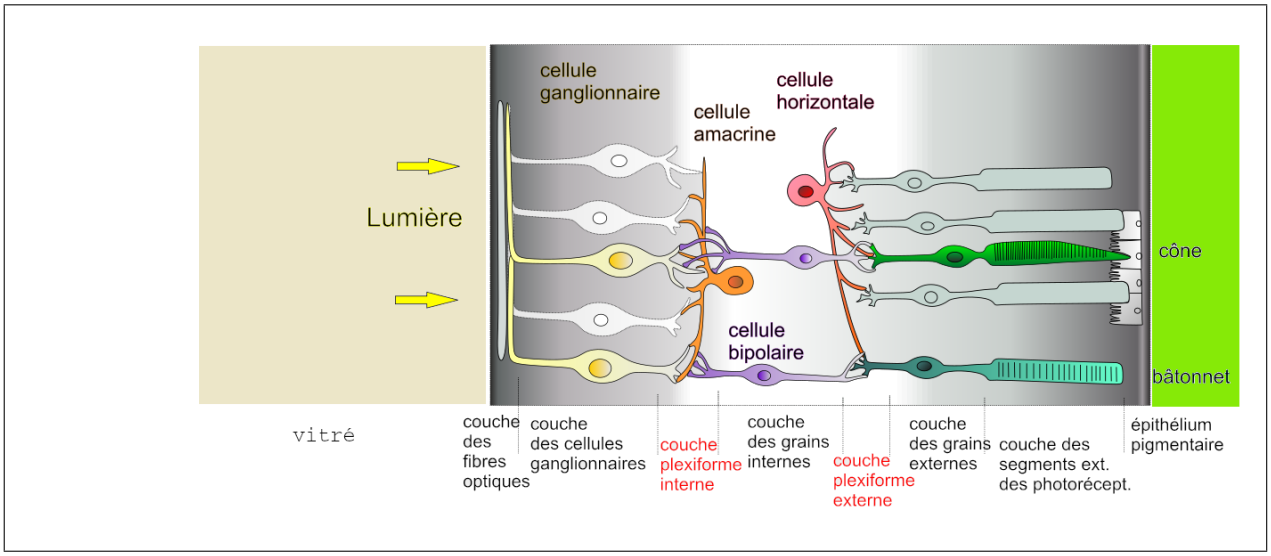

Figure 1.5 – *Vue en coupe schématique de la rétine. La lumière provient de la gauche après avoit traversé la cornée, le cristallin et le vitré. Contrairement à l'intuition elle doit traverser une épaisse couche de cellules avant d'être absorbée par les pigments qui se trouvent à l'extrémité des cônes ou des bâtonnets, c'est-à-dire orientés vers l'intérieur de la tête - (©Wikipedia).*

synaptique C'est le début d'une longue chaîne de traitements accomplis tout d'abord dans la rétine elle-même, puis dans le nerf optique et enfin dans les aires visuelles. Cette chaîne se fait de façon électrique au sein des neurones et de façon chimique entre les neurones aux synapses.

Dans la rétine (voir figure 1.5), ces traitements sont initiés par des connexions entre cellules voisines ou proches. Ce sont successivement :

- les cellules horizontales qui relient les sorties des cônes ou des bâtonnets dans des liaisons tangentielles (donc entre récepteurs voisins), formant la couche plexiforme externe,
- les cellules bipolaires qui relient la couche plexiforme externe à la couche plexiforme interne, et réalisent des opérations élémentaires de traitement du signal optique (augmentation de contraste par addition ou détection par soustraction et amplification),
- les cellules amacrines qui, comme les cellules horizontales ont un rôle de transmission tangentielle à des cellules ganglionnaire voisines, des informations issues des cellules bipolaires. Elles construisent la couche plexiforme interne,
- les cellules ganglionnaires dont les axones se regroupent pour former le nerf optique.

Le nerf optique ainsi constitué doit retraverser la rétine en raison de l'orientation inversée des cellules. Cela se fait en un lieu appelé papille (ou tache aveugle car il n'y a aucun capteur en ce lieu) située un peu hors de l'axe optique de l'œil au repos [6].

Les cellules échangent des informations au niveau de leurs synapses, sous forme d'impulsions temporelles. Une même cellule peut être en relation avec plusieurs centaines de cellules réceptrices. Ainsi, des cellules bipolaires peuvent sommer les contributions d'un millier de bâtonnets, ce qui explique la grande sensibilité de la vision nocturne, mais aussi sa faible résolution spatiale. Certaines cellules sont stimulées par la lumière, d'autres sont inhibées. La connexion de cellules inhibées et de cellules stimulées permet de construire des détecteurs (cellules ON-OFF) (figure 1.6). Les champs récepteurs des zones ON et OFF pouvant couvrir des distances différentes, les détecteurs sont ainsi sensibles à des détails plus ou moins fins [7]. Des groupements sont plutôt sensibles aux variations temporelles,

6. Il est remarquable que la plupart d'entre nous ne seront jamais conscients de cette zone aveugle que l'on ne découvre que par des petites expériences personnelles (faire disparaitre l'image d'un petit objet en le déplaçant dans le champ visuel en vision monoscopique). La vision spontanée permet de parfaitement compenser ce défaut.

7. Les traiteurs d'image se sont inspirés de ces schémas pour concevoir des détecteurs de contours, par exemple ceux de Canny [**?**] ou de Deriche.

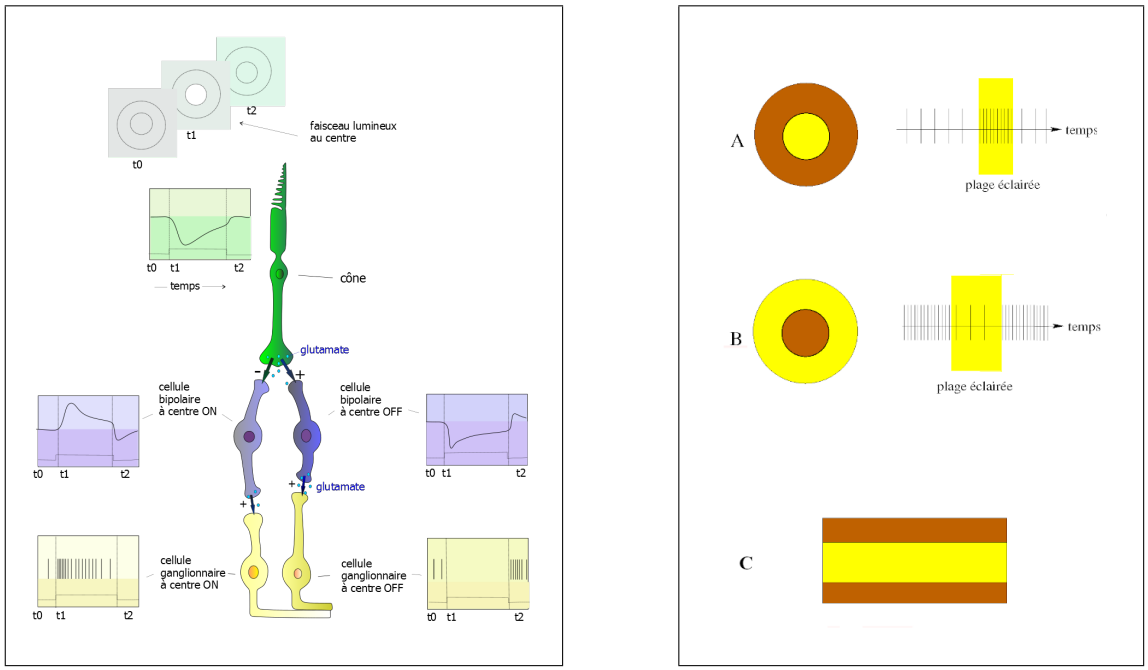

Figure 1.6 – *A gauche, construction de 2 cellules ON-OFF à partir d'un photorécepteur. La cellule bipolaire à gauche a un fonctionnement ON (elle est active en présence de lumière), celle de droite a un fonctionnement OFF (©Wikipedia). A droite, en A, cellule ganglionnaire ON-OFF de la rétine éclairée en son centre : la fréquence de la réponse temporelle de son potentiel d'action augmente lors de l'éclairement. En B c'est une cellule OFF-ON. En C, champ récepteur ON-OFF allongé d'une cellule des aires visuelles propre à amplifier les contours.*

d'autres à des répartitions de motifs (des textures), d'autres sont spécialisés dans les contrastes chromatiques. Ainsi sont élaborés, dès la rétine, une vingtaine de flux d'informations différents, qui sont transmis simultanément, ce qui explique bien certaines propriétés de la vision humaine : une grande capacité à détecter des mouvements fins et des faibles variations de luminosité (surtout en périphérie du champ visuel en raison de la présence de nombreux bâtonnets), une grande sensibilité au contraste et aux contours, une capacité à identifier en une seule zone des textures semblables, une sensibilité aux nuances chromatiques (diversité des pigments, associations des cellules horizontales).

Les cellules réceptrices étant au nombre de cent millions environ, et les fibres du nerf optique de l'ordre d'un million, on voit l'importance de ces connexions intermédiaires qui mettent progressivement en forme le signal visuel et constituent un codage sophistiqué de l'information.

1.1.3 Les voies optiques

Les signaux élaborés dans la rétine sont transférés par les nerfs optiques aux aires visuelles situées dans la partie postérieure du crâne (figure 1.7 à droite), ([Purves et al., 2015], chap. 12).

Les deux parties du champ visuel de chaque œil, la partie temporale et la partie nasale, sont séparées et transmises séparément : la partie temporale est traîtée par les aires situées du même côté (les aires droites traîtent les signaux temporaux issus de l'œil droit) tandis que les parties nasales sont échangées lors du passage à travers le chiasme optique. Cet échange des informations issues d'un même point de l'espace mais vues par chacun des deux yeux (figure 1.7), contribuera à la reconstitution stéréoscopique du relief, que nous verrons plus loin (paragraphe 1.4). A la suite du chiasme, les informations sont traîtées de façon contraposée : la partie gauche des aires visuelles traîte prioritairement les signaux de l'œil droit et vice-versa.

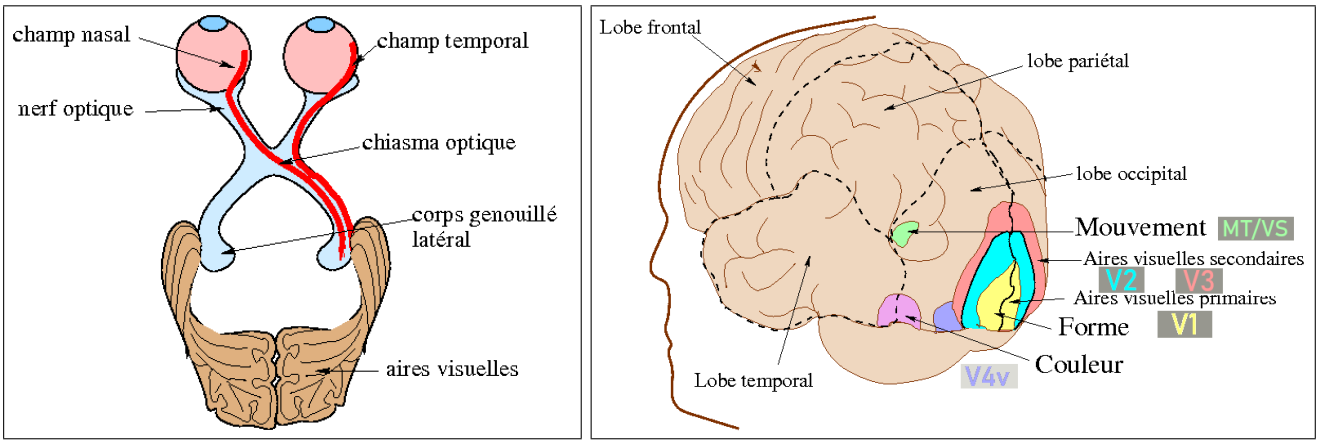

Figure 1.7 – *à gauche : chiasme optique et regroupement des signaux des 2 yeux issus d'un même point de la scène de façon à pouvoir faire traiter la disparité stéréoscopique par une même aire visuelle. à droite : les aires de la vision sont essentiellement regroupées dans la partie occipitale du crâne. La propagation du signal se fait de la base occipitale vers les lobes pariétaux et temporaux et l'interprétation s'élabore au cours de cette progression.*

Après le chiasme optique, un million de fibres environ conduisent le signal visuel par le nerf optique vers les corps géniculés latéraux (ou corps genouillés) où il subit un très important traitement. Le corps genouillé met en forme le signal visuel à destination des aires visuelles V1, V2, V3 et V4 [8], mais il agit en étroite interaction avec ces aires, car il existe une boucle de contre-réaction entre V1 et le corps genouillé. Au sein du corps genouillé, le signal visuel est réparti en trois voies visuelles aux attributions bien identifiées. Les deux plus importantes : la voie parvocellulaire (voie P) et la voie magnocellulaire (voie M) sont construites sur les signaux de contraste lumineux issus des récepteurs L et M, transmettant une dominante chromatique Jaune-Vert. La voie K (koniocellulaire), construite à partir des cellules S, transmet une image des Bleus.

La voie P est tonique (elle transmet un signal tant que le stimulus est présent), elle est composée de cellules de petite taille, transmet des informations sur des stimulus de faible étendue, à variation spatiale élevée mais à lente évolution temporelle. Elle sera très importante pour la détection statique, pour la reconnaissance et pour l'analyse fine de la scène, par exemple pour la lecture.

La voie M au contraire est phasique (le signal n'existe qu'à l'apparition ou à la disparition du stimulus), elle est constituée de cellules très grandes, à faible résolution spatiale, mais à très bonne résolution temporaire. Elle prendra en charge les informations de mouvement et en particulier la détection des événements brefs.

Le voisinage des cellules de la rétine qui ont donné naissance à l'image est maintenu dans les voies optiques et transporté jusqu'au corps genouillé latéral d'où il est transmis aux aires visuelles. Cela permet de transmettre une carte de l'espace 3D observé (appelée **carte rétinotopique**) qui facilitera l'interprétation des signaux visuels. Sur la carte rétinotopique les propriétés topologiques de la scène observée sont respectées, mais pas ses propriétés métriques. La zone observée par la fovéa occupe une place très importante, tandis que les régions en bordure de champ sont progressivement réduites selon une projection homéomorphe [9]. Ainsi les aires visuelles sont alimentées par un flot d'informations déjà très fortement organisé et filtré (en termes de contrastes chromatiques, de fréquences spatiales et d'orientation des stimulus), mais avec une topologie totalement identique à celle du stimulus visuel

8. Le flot principal du signal est dirigé vers les aires visuelles (c'est de lui que nous parlons dans la suite), mais une partie du signal est également renvoyée vers des aires du cortex moteur pour servir principalement au contrôle des muscles en charge de la vision : fixation, accomodation, convergence, micro-saccades et éventuellement pour permettre des actions réflexes de recul, d'évitement ou de protection.

9. Divers modèles mathématiques ont été proposés pour rendre compte de cette projection : modèle "monopôle" $= \log(z+a)$, modèle "dipôle", modèle "log d'une homographie", ... Ces transformations sont cependant très difficiles à mesurer en raison de la surface très complexe du cerveau [Petitot, 2008].

reçu.

1.1.4 Les aires visuelles

Les premiers niveaux que nous venons d'examiner ont permis de faire une analyse très locale fine et pertinente du signal visuel. Certaines aires du cortex, situées dans la partie occipitale du crâne, sont chargées de conduire une analyse beaucoup plus structurée et plus complète de ces informations locales, de les replacer dans un contexte plus large (tirant profit de sa proximité dans la carte rétinotopique) et de conduire à une analyse plus vaste indispensable à la compréhension de la scène ([Purves et al., 2015], chap. 12).

Ce sont les aires visuelles primaires et secondaires (les aires V1 à V4) qui se spécialisent dans le traitement systématique de certains types d'information : la forme, la couleur, les contours, le mouvement. Les traitements sont entrepris d'une façon parallèle selon un schéma fixe de "colonnes" spécialisées. Pour une région de l'espace visuel, l'information liée à un flux de propriétés particulières (texture, contraste chromatique ou orientation des contours), les signaux se propagent dans une colonne qui maintient la localisation des informations, une autre colonne voisine étant en charge par exemple d'une autre orientation. Ces colonnes sont regroupées en hyper-colonnes. Les cellules réceptrices des aires primaires ont des champs d'activation plutôt allongés (avec également une zone activatrice et une zone inhibitrice, comme les cellules de la rétine) (voir figure 1.6). Cette disposition facilite les opération d'intégration spatiale le long des contours.

Des expériences historiques, comme celles de Hubel et Wiesel [Hubel and Wiesel, 1959] dans les années 1950-1960, ont permis de vérifier que ces aires effectuent des transformations très proches de l'analyse de Fourier ou de l'analyse en ondelettes du champ visuel en distinguant les orientations dans le plan ainsi que les fréquences spatiales selon des échelles d'étendues différentes [10]. Ces traitements s'appliquent aussi dans l'espace tridimensionnel tel que permet de le reconstruire la stéréovision opérée à partir des comparaisons des champs droit et gauche superposés après traversée du chiasma optique. Les décompositions fréquentielles et directionnelles sont directement utilisées pour la détection des formes et la reconnaissance des objets ; combinées avec les informations temporelles, elles permettent de réaliser de façon purement réflexe des opérations complexes de surveillance (détection d'un mouvement singulier dans une scène en mouvement par exemple).

1.1.5 ... et ensuite ?

A partir de ce niveau de traitement, nos connaissances sont beaucoup moins sûres de la façon dont les diverses cartes sont conjointement exploitées, de la nature des signaux qui sont transmis aux aires suivantes, ainsi même que des aires qui sont en charge des traitements.

Car le cheminement des signaux issus de la rétine n'est pas achevé. A partir des aires visuelles, les représentations sensorielles, prenant en compte les formes, les couleurs, les textures, le mouvement, la distance et le relief sont transformées en représentations cognitives. D'autres aires du cerveau sont alors sollicitées, en particulier les aires du cortex liées à la mémoire (hippocampe), au langage (aire de Broca), éventuellement à la lecture (cortex occipito-temporal gauche), les aires en charge des émotions (amygdales, noyaux postérieurs du thalamus) ainsi que des aires associées à la récompense (téléencéphale) si l'image est plaisante ou au contraire désagréable, les aires associées à la préparation d'une action (lobe frontal et ganglions de la base) : fuite, parade, mouvement, si la scène présente des éléments de danger. En parallèle, comme nous l'avons vu dans le chiasme, des signaux visuels sont adressés à des aires spécialisées dans le contrôle en boucle de réaction du système visuel, pour assurer une mise au point, suivre une cible, explorer un contexte, confirmer une interprétation par la focalisation du regard sur un détail, solliciter une expertise particulière (estimation d'une taille, d'une distance, recherche d'un critère de désambiguation) afin de mettre en œuvre des processus

10. Le traiteur d'image saura reproduire ces mécanismes à l'aide de décomposition pyramidale ([Adelson et al., 1984] incorporées formellement dans la *scale-space theory* de T. Lindeberg [Lindeberg, 1994].

spécialisés de reconnaissance (lecture, orientation, ...), aider à l'interprétation de l'émotion d'un visage, à la prévention d'un danger ou d'une évolution d'une situation, à la réussite d'un mouvement de préhension, etc. Il est probable que ces boucles de retour ne soient pas faites à partir des conclusions immédiates des aires visuelles, mais à la demande d'autres régions du cortex, elles participeraient ce que l'on appelle la voie limbique ou sous-corticale.

Nous prendrons le seul exemple de la reconnaissance d'un texte écrit. Si l'on sait que les aires visuelles analysent chaque caractère du texte en termes de longueur, largeur, orientation, courbure, boucles et composantes connexes, c'est *l'aire de la forme visuelle des mots* située sur la tempe gauche, un peu en arrière de l'oreille, qui utilise tout ce matériel pour composer des mots indépendemment des lignes précises qui le composent (taille, police, support). Cette analyse est détaillée dans [Gaillard, 2024] et est présentée en figure 1.8. Le même type de compilation et intégration des primitives de base fournies par les aires visuelles, est assuré par une aire de reconnaissance des visages située assez symétriquement dans une aire occipito-temporale droite ([Gaillard, 2024]).

Mais ce sont les étapes suivantes de la lecture qui s'avèrent plus passionnantes encore. Votre cerveau se livre à un travail de décodage lui permettant de reconnaître les lettres que forment ces traits, puis les syllabes qu'elles forment entre elles : c'est le b.a-ba de la lecture. C'est le long du cortex occipito-temporal gauche, c'est-à-dire derrière votre oreille gauche, que se produit cet assemblage. En quelques centaines de millisecondes, il conduit à l'activation d'une région experte de notre cerveau, l'aire de la forme visuelle des mots. Celle-ci porte mal son nom, puisqu'elle est largement indépendante de cette forme visuelle, ne faisant pas la différence entre ce MOT ou ce mot, c'est-à-dire codant l'identité de ce mot indépendamment de son écriture. Songez que les deux objets visuels que constituent ces deux écritures du même mot sont bien différents l'un de l'autre, et que cette aire cérébrale contient donc un code relativement abstrait de ce mot, en tant qu'assemblage spécifique de lettres quelle que soit la forme de ces lettres. À l'inverse cette aire différencie bien les mots HACHE et VACHE, dont la forme visuelle est extrêmement proche, mais qui sont assurément des mots différents. Il faut donc que votre cerveau soit capable d'associer entre elles les lettres HACHE, même si nous en espaçons les lettres H A C H E au point d'ouvrir – à la hache – une béance entre chaque lettre que votre cerveau tout aussitôt referme[1]. Certes, certaines façons de l'écrire viendront ralentir votre lecture, ainsi en est-il de cette HACHE, et même si le gothique sied parfaitement à l'objet, mais eNtRe mInUsCuLEs et MAjuSCUles, IL voUS ESt PerRmIS dE LiRe SAnS difFIcULltéS ceTTe pHRaSe quasiment sans entraînement et malgré des différences de forme majeure. Il est convenu de parler d'*invariance de forme*, votre cerveau étant dans l'ensemble capable de coder la lecture d'un mot indépendamment de la forme de ce mot.

Figure 1.8 – *De la perception des caractères à la lecture d'un texte. Extrait de l'ouvrage* L'homme augmenté *pages 28-29 [Gaillard, 2024] : Après la décomposition de l'image en une batterie de primitives locales par les aires occipitales du système visuel, la reconnaissance des mots est assurée par une aire spécialisée du cortex occipito-temporal gauche, indépendamment des formes précises du texte particulier présenté.*

Nous voudrions, pour convaincre lecteur de la complexité des ressources mobilisées pour des tâches élémentaires de vision, évoquer des expériences vécues par chacun. Se diriger au sein d'une foule dense ou sur une piste de ski passe par la détection de certains indices favorables au déplacement et nous fait ignorer la plupart des composantes de la scène qui n'interféreront pas avec le mouvement (... ou du moins le croît-on !). Lire un texte ou une partition, pour une personne expérimentée, c'est ignorer beaucoup des symboles pour ne retenir que ceux qui accompagnent une interprétation élaborée simultanément à la perception. Ce sont des exemples vécus par chacun qui montrent que l'acte de perception n'est pas le seul résultat de la transmission d'un message visuel, mais qu'il peut faire appel à des circuits très différents en fonction de l'éducation et de la culture de l'observateur. Rechercher une photo dans une collection (chacun le fait sur son téléphone ou son ordinateur), c'est passer très rapidement sur des dizaines de documents qui ne correspondent pas à une esquisse que nous nous sommes mentalement fixée. Clairement, la tâche de recherche fait plus appel à notre capacité à identifier une image conforme à un modèle préexistant qu'à analyser et bâtir une interprétation. Reconnaître un ami dans une photographie ou à la gare lors d'un rendez-vous : quelle part de vision "passive" ? quel rôle pour la mémoire ? est-ce que l'on corrèle des signaux ? est-ce que l'on déforme

des souvenirs pour les fixer sur une image ? Beaucoup d'ignorance demeure dans la compréhension de nos actes les plus quotidiens mais cette ignorance ne semble pas nuire à l'efficacité du processus.

Nous verrons un peu plus loin, dans la partie de ce texte consacrée aux "illusions d'optique", comment on peut vérifier de façon reproductive le rôle de certains mécanismes mis en jeu dans notre perception (section 1.3).

L'enchaînement de ces étapes à partir de la capture des photons par les photorécepteurs de la rétine jusqu'à l'interprétation de la scène observée montre bien comment la vision, initialement processus physiologique, passif et objectif d'enregistrement d'un signal extérieur à l'observateur, devient progressivement un processus cognitif, actif, mobilisant les aires supérieures du cortex, dans une interaction intime avec notre acquis conscient et inconscient [Meyer, 1997], donc un processus psychologique. C'est cette dualité physiologie/psychologie qui rend compte de la complexité des mécanisme de la vision. Nous le disions, l'enchaînement des opérations nécessaires à l'accomplissement de ces tâches est encore largement méconnu et débattu. L'existence et le rôle de la voie sous-corticale sont questionnés [Pessoa and Adolphs, 2010] ; des travaux récents suggèrent que les aires visuelles V1 et V4, procèdent, très en amont de l'amygdale et du thalamus, à une détection des émotions [Kragel et al., 2019], mais cela est débattu. Nous savons, grâce aux techniques d'imagerie (surtout l'imagerie par résonnance magnétique fonctionnelle (IRMf)), quelles aires de notre cerveau sont actives lors de ces diverses étapes, mais nous ne connaissons ni l'ordre de leur activation, ni les mécanismes de leur intervention car nos outils actuels sont trop lents et trop incertains [11].

Selon une doctrine très répandue, il semblerait aujourd'hui que l'on doive combiner deux circuits différents de l'information. L'un serait rapide et conduirait à une évaluation quasi immédiate de la situation à partir du seul flot issu des aires visuelles, l'autre prendrait quelques secondes et mobiliserait les aires plus cognitives du cortex cérébral [Damasio, 1994a, Damasio, 1994b]. Peut-être ce second circuit pourrait-il être activé précocement à partir de signaux qui lui seraient destinés dès le corps genouillé, mais cela est encore l'objet de spéculations. Dans d'autres modèles les flux issus des divers traitements sont dirigés chacun vers les aires en charge de l'action (par exemple le mouvement, la reconnaissance, l'émotion) puis créent un bain d'impressions, toutes en concurrence, qui va contribuer à former notre « conscience » de la scène qui nous entoure [Dennett, 1993].

Cette complexité de la vision et son intrication avec la conscience que nous avons du monde qui nous entoure est aujourd'hui bien exprimée dans les représentations phénoménologiques [12] adoptées par une grande partie de la communauté philosophique où l'on ne sépare plus, dans un objet précis de l'univers (une pomme, un chien, un ami, ...) les aspects issus de la perception (ce que nos yeux nous transmettent) de ce que nous en « savons ». Conscience procédurale (ou psychologique) qui contrôle et commande nos actions, conscience phénoménologique qui gère notre capacité à appréhender ce que nous ressentons à un moment donné. La première qui serait *facile* pour une IA, tandis que la seconde serait le vrai défi pour les informaticiens : l'IA *difficile* [Chalmers, 2010].

11. L'outil majeur d'analyse fonctionnelle du cerveau est l'imagerie par résonnance magnétique fonctionnelle (IRMf) qui fournit une excellente résolution géométrique (quelques millimètres), et une très bonne localisation tridimensionnelle [Gori, 2018, Brown et al., 2007]. Mais elle contraint à un protocole très lourd. Par ailleurs le signal qu'elle délivre est très faible et la résolution temporelle est médiocre (plusieurs secondes). Pour exploiter le signal, il faut généralement moyenner les résultats d'une dizaine d'expériences. Si ces résultats proviennent de plusieurs sujets, il faut procéder à une projection des signaux sur un même atlas universel. Parmi les autres modalités, l'électroencéphalographie (EEG), offre des protocoles d'étude plus simples. Les mesures obtenues concernent surtout les aires externes du cerveau. Elles ne sont pas très bien résolues spatialement (un ou deux centimètres) mais ont une excellente résolution temporelle (quelques dizaines de millisecondes) permettant de distinguer les détections par les aires primaires (100 à 150 ms) des opérations plus complexes de catégorisation (350 à 500 ms) ou encore des opérations lentes (1 s ou plus) mettant en jeu la mémoire ou les émotions, [Thorpe et al., 1996, Fize, 2004, Schupp et al., 2004]. Enfin les études par mesure électrique directe au sein du cerveau qui ont le mérite d'être à la fois précises et rapides sont réservées à des situations exceptionnelles en raison de leur caractère invasif.

12. La représentation phénoménologique de Milou prend en compte la silhouette familière et blanche, au contour bouclé, à la queue courte dressée, que nous a transmis notre lecture des albums (la composante perceptive) mais aussi la notion d'un chien turbulent, fidèle à son maître, alternativement téméraire et peureux, un chien qui ne mort que les méchants ... (les aspects cognitifs, discutables, douteux peut-être). Lorsque je « vois »Milou, c'est ce contexte qui accompagne une nouvelle vignette.

1.2 Quelques propriétés de la vision

1.2.1 L'œil et la caméra

Revenons sur certaines propriétés du système visuel et comparons les aux propriétés des capteurs artificiels : appareil photo ou caméra.

En ce qui concerne la résolution tout d'abord, rappelons que le champ visuel est très inégalement couvert. Seule la partie centrale de l'image, saisie par la fovéa bénéficie de la haute résolution que nous avons évoquée : une minute d'arc [13]. Nous nous rendons très peu compte de l'étroitesse de ce champ en raison de la grande mobilité de l'œil qui nous permet d'explorer très rapidement un champ beaucoup plus large en nous arrêtant à haute résolution sur toutes les plages qui retiennent notre attention. Mais il ne faut pas négliger l'importance de ces zones périphériques dans notre compréhension d'une scène car ce sont les signaux issus de ces zones qui président au choix des régions que nous explorons de façon réflexe par les saccades (figure 1.2).

Il est difficile de comparer ces performances en résolution avec celles d'un appareil photographique dont le capteur est homogène et ne dispose pas de zones de focalisation. On ne sait pas aujourd'hui, à des coûts raisonnables, couvrir le champ visuel de l'homme avec la résolution ultime de la fovéa, cependant la plupart des capteurs délivrent plus de pixels utiles que la fovéa, mais de façon systématique, ce qui conduit à des calculs nombreux, souvent inutiles. En contrepartie, l'œil est remarquablement agile dans ses déplacements et sa focalisation, ce qui permet de tirer le meilleur bénéfice de ses capacités.

L'œil ne perçoit que les rayonnements de la bande comprise entre 400 nm et 750 nm, tandis que les caméras peuvent être poussées beaucoup plus loin dans l'infra-rouge. L'œil perçoit 3 canaux LMS dans le domaine visible, ce que la caméra fait assez bien à l'aide des matrices de Bayer RGB, mais les caméras peuvent aussi fournir des données multispectrales à condition de perdre en résolution.

La courbe de sensibilité de l'œil varie beaucoup en situation diurne ou nocturne ainsi qu'en fonction des longueurs d'onde incidentes. Les courbes de réponse des cônes et des bâtonnets sont rapportées sur la figure 1.3. Comme toutes les réponses physiologiques, les cônes et les bâtonnets ont un comportement logarithmique en fonction de l'énergie incidente (loi de Fechner), bien adaptée aux fortes dynamiques. La vision du jour, dite photopique, assurée par les seuls cônes, peut couvrir une gamme de luminances allant de 10^{-3} à 10^{8} cd/m^2, ce qui correspond à une nuit de pleine lune et à un plein soleil d'été [14] . Les bâtonnets prennent le relai pour les très faibles intensités de 1 à 10^{-6} cd/m^2. La perception des couleurs disparaît vers 10^{-2} ou 10^{-3} cd/m^2, début d'une zone de vision dite scotopique, séparée de la région photopique par une zone intermédiaire dite mésopique où les cônes et les bâtonnets sont simultanément actifs (Figure 1.9). Le passage de la vision photopique à la vision scotopique nécessite un temps long (plusieurs minutes, voire plusieurs dizaines de minutes).

Comparant ces performances à celles d'une caméra, on constate que peu de capteurs numériques disposent d'une si grande dynamique, mais par combinaison des divers réglages d'une caméra (le gain en sensibilité ISO, l'ouverture du diaphragme, le temps d'exposition), ou par association de divers capteurs, on parvient à des dynamiques supérieures avec les meilleures caméras. En particulier, on conçoit aujourd'hui des caméras qui, soit pour les très fortes luminosités, soit pour les très faibles, excèdent les performances de l'oeil.

L'œil ne dispose pas d'une très grande rapidité de réponse. La partie périphérique de la rétine permet de rendre compte d'événements très rapides (le millième de seconde environ) s'ils sont lumineux (étoile filante, éclair), mais l'image qui en est donnée est médiocre. La fovéa, à pleine résolution, peine à rendre compte de phénomènes plus brefs que le dixième de seconde [15]. En comparaison, la

13. En fait, comme nous l'avons vu, l'œil souffre même d'une zone aveugle (la papille, voir figure 1.1). Ce défaut est parfaitement surmonté par notre stratégie de perception qui vient combler cette absence de signal sur un œil par l'information correspondante collectée sur l'autre œil.

14. Rappelons cependant que si les cônes ont une large plage d'excursion en sensibilité, il leur faut plusieurs dizaines de secondes pour s'adapter à des changements rapides, la variation de taille de l'iris n'agissant que de façon assez marginale sur le flux de lumière reçu par la rétine.

15. Il est frappant de se rappeler que l'on n'avait pas une connaissance exacte de la position des pattes d'un cheval

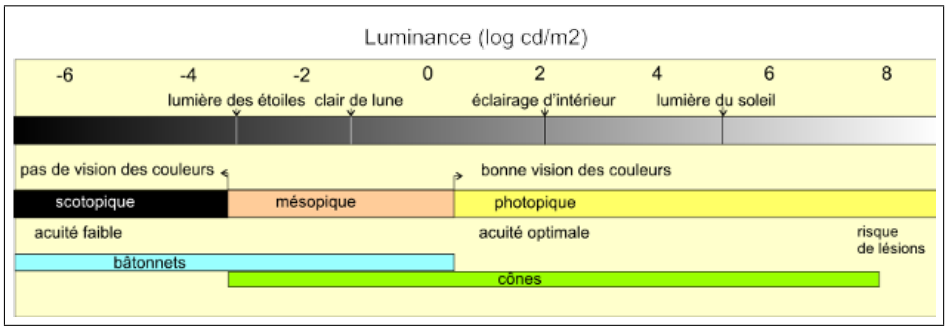

Figure 1.9 – *Gamme des luminances auxquelles fonctionne le système visuel. Les 3 termes "scotopique, photopique et mésopique" désignent les gammes d'énergie pour lesquels les bâtonnets fonctionnent seuls (scotopique), les cônes fonctionnent seuls(photopique), ou les deux types de cellules sont simultanément actives (Mmsopique). L'échelle des énergies lumineuses est ici en log_{10} candela/m^2. (© Wikipedia)*

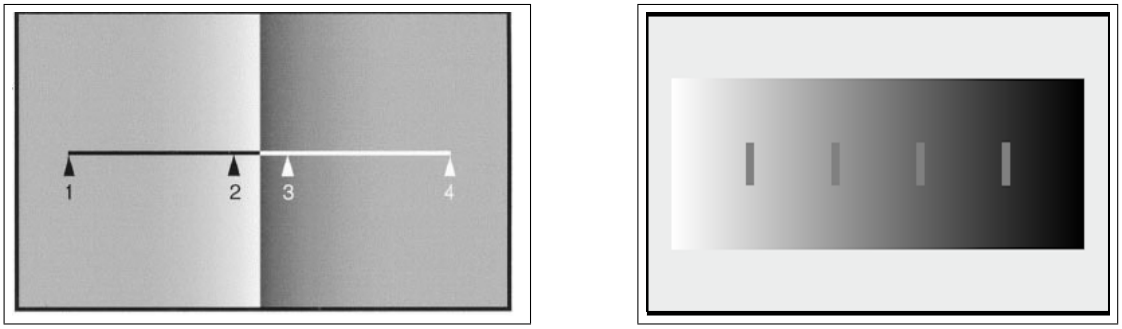

Figure 1.10 – *Deux exemples de fonctionnement actif du système visuel. Ces phénomènes peuvent s'expliquer à l'aide des propriétés des traitement des cellules visuelles dès la sortie de la rétine. A gauche une bande de Mach (c'est-à-dire un renforcement local du contraste) apparaît au voisinage d'un fort contraste. La luminance est plus claire près de 2 qu'en 1 et plus sombre en 3 qu'en 4. A droite, le contraste simultané est un exemple classique de l'interaction entre le fond et l'objet d'avant-plan : les quatre barres ont le même niveau de gris mais sont perçues de façon très différente selon le fond sur lequel elles se trouvent.*

photographie permet des expositions très inférieures au millième de seconde

1.2.2 Perception passive versus perception active

Ce qui différencie fondamentalement l'image fournie par l'œil de celle que donnerait une caméra est le rôle important des interactions spatiales entre cellules voisines. Ainsi, la perception que l'on a d'une plage de gris dépend-elle des niveaux de gris adjacents (comme on peut le vérifier sur la figure 1.10 à droite) et trompe notre perception (comme dans le cas des bandes de Mach, figure 1.10 à gauche). Ces interactions spatiales, en amplifiant les signaux ou même en créant des signaux inexistants, comme nous le verrons, facilitent (ou préparent) la détection des objets. Ce rôle actif se manifeste également pour les variations lentes et dégradées qui sont gommées, pour les ombres qui sont très fortement réduites. Plus complexes encore sont les interactions colorées qui, non seulement modifient les contrastes mais également déplacent les couleurs perçues dans un espace perceptuel subjectif. Ces propriétés ont été mises à profit par les peintres depuis plus d'un siècle (voir par exemple les travaux de Seurat ou, plus proche de nous, de Vasarely).

Les capacités de l'œil à accentuer les contrastes, à gommer certains détails, à faire abstraction de variations dues à l'éclairage ou au masquage ou encore la distorsion géométrique sont le propre d'un capteur **actif** qui prépare le signal à des tâches vitales d'interprétation en vue d'une action

au galop avant les expériences de ciné-photographie de Marey et Muybridge.

(fuite, protection, préhension, ...). De multiples prototypes de caméras ont été proposés pour copier ces facultés (les *smart cameras*), avec des résultats plus ou moins heureux. Outre les difficultés d'intégration de ces fonctions dans les rétines artificielles, se pose le problème de savoir jusqu'à quel niveau il est souhaitable d'intégrer les fonctions connues chez l'homme. Plus difficile encore est la conception d'une rétro-action qui existe dans le système visuel mais que l'on peine à concevoir dans la machine. Enfin la vision humaine est non seulement active, mais aussi **subjective** en ce qu'elle mêle des éléments conscients et inconscients à l'interprétation de l'image. Cela rend la simulation beaucoup plus complexe.

1.2.3 Une vision subjective

Montrons dans un exemple simple (figure 1.11) comment notre façon de penser le monde peut s'imposer à la vision que nous en avons. L'image semble relever des mêmes mécanismes que celle que nous avons décrite en figure 1.10 à droite où l'apparence de teintes identiques dépend du contexte des plages (le gris perçu est influencé par les teintes qui l'entourent).

Cependant, sur cette image 1.11 cette interprétation simple est insuffisante. En B, les quatre plages grises sont perçues plus claires que les trois plages grises en A (pourtant d'un même gris). Or elles sont en B sur un fond majoritairement plus clair qu'en A et devraient à ce titre être perçues plus sombres. Que se passe-t-il ? L'explication communément admise de ce cas délicat [Purves et al., 2015] fait appel à une interprétation complexe et tridimensionnelle de la scène qui n'attribue que les plages claires pour seuls voisins des plages considérées en A, et que les plages noires, en B. Ainsi les plages grises en A sont assombries et celles en B sont éclaircies, en imaginant qu'elle proviennent de bandes qui seraient "tissée" et dont le seul voisinage visible est celui des parties qu'elles cachent : claires en A, sombres en B. Notre "perception" est bien ici le résultat d'une "interprétation" (inconsciente et réflexe le plus souvent) qui essaie de "comprendre" cette image en la rapportant probablement à d'autres scènes semblables déjà perçues. La voie optique vient se plier à une image mentale. La vision est subordonnée à l'interprétation. C'est cette même explication qui est donnée pour expliquer l'illusion de l'échiquier d'Adelson (figure 1.15). Dans ce cas l'image inconsciente, mais universellement partagée, qu'un échiquier est formé de cases identiques donne à l'argument une force supplémentaire.

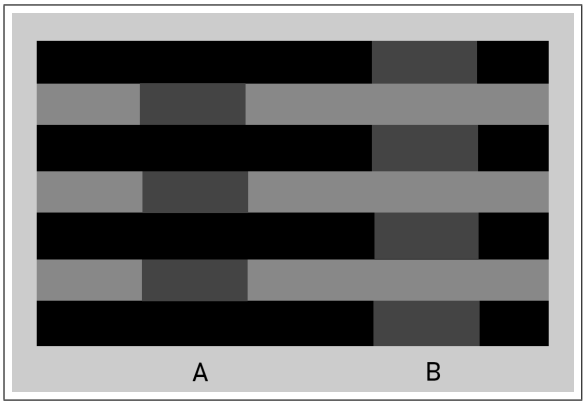

Figure 1.11 – *La perception différente des 3 zones grises en A et des 4 zones grises en B à droite ne peut pas s'expliquer sans le recours à des mécanismes d'interprétation de l'image. En effet, en A les plages grises sont entourées majoritairement de gris sombre et devraient apparaître donc plus claires si l'on s'en tenait aux explications des figures 1.10. Si ce n'est pas le cas, c'est que l'on interprète la bande verticale en A comme appliquée sur le fond clair, tandis qu'en B on la voit sur le masque noir.*

La vision, processus actif, peut non seulement modifier les teintes et les formes perçues, mais aussi introduire des éléments étrangers à l'image, comme les faux contours que l'on observe dans la figure 1.13. Ces phénomènes sont le résultat de deux traitements différents qui se combinent : un traitement physiologique, s'appuyant sur les connexions entre cellules (assez bien modélisé par le schéma actif

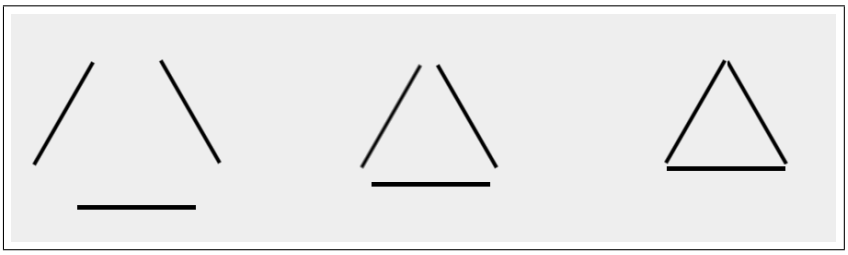

Figure 1.12 – *Le triangle est vu comme une unité graphique en soi lorsque les trois traits se rapprochent.*

décrit ci-dessus), un traitement psychologique (celui que l'on a dit "subjectif") qui fait appel à des aires supérieures de la vision et probablement d'autres aires du cortex et qui emprunte les circuits de rétro-action dont nous avons parlé plus haut. Il n'existe pas de théorie complète modélisant précisément et mathématiquement l'ensemble de ces processus, mais de nombreux éléments de réponses qui se raccordent mal et qui parfois se contredisent. Nous allons présenter ici deux modèles très différents de la vision. Nous commencerons par le plus récent qui formalise les premiers étages de la vision : la théorie computationnelle de Marr (section 1.2.5), puis nous examinerons une théorie de psycho-physiologie, plus ancienne, étayée par des expérimentations depuis un siècle, qui a été laissée de côté pendant de nombreuses années, mais revient à l'avant-scène : la théorie de la Gestalt (section 1.2.6). L'une des approches essaie de fournir un modèle qui part du stimulus physique, l'autre, au contraire, s'appuie sur la conclusion subjectivement éprouvée en fin de processus perceptuel.

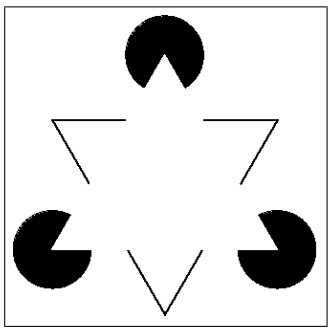

Figure 1.13 – *Triangles subjectifs de Kanizsa. Cette figure, si l'on veut la décrire exactement par de seuls éléments tracés, conduisent à une description complexe : "*3 portions de cercles disposés en triangles, dont les secteurs absents sont orientés vers le centre du triangle, etc.*". Une interprétation semble plus simple :" un triangle blanc unique recouvre un triangle noir et 3 cercles". Il semble que ce soit cette interprétation qu'adopte notre perception. Afin de conforter cette interprétation l'observateur "perçoit" subjectivement les bordures manquantes du triangle "supérieur" et induit une différence de teinte entre l'intérieur et l'extérieur de ce triangle virtuel, différence qui n'existe pas. (d'après G. Kanizsa, Scientific American, avril 76).*

Mais avant de quitter la description générale du phénomène de vision et de perception, il est souhaitable de dire un mot de la situation extrême de la perception d'images sans la mise en œuvre du système visuel.

1.2.4 Des images sans vision

La capacité de l'homme de "voir" des images sans mettre en œuvre l'œil et la perception est connue de chacun de nous à travers ses rêves, voire des expérience hallucinogènes dues à la maladie ou aux drogues. Plus élémentairement, nous pouvons simplement en faire l'expérience en fermant les yeux, faisant ainsi apparaître des taches, des contrastes, des couleurs, mouvants dans le temps. Une absence de "signal" sur la rétine peut donc s'accompagner d'une perception visuelle. Celle-ci est très liée au dernier signal perçu (inversion de contraste, couleurs complémentaires) exprimant le retour

au repos de la chaîne électrique et chimique des voies visuelles. Mais au bout d'un moment, cette perception s'organise selon des motifs qu'il est très difficile d'interpréter mais qui ne semblent liés ni aux images vues, ni aux structures mentales conscientes. A l'inverse, les rêves nous ramènent à nos expériences visuelles anciennes parfois très lointaines et parfois avec une grande précision (visages, lieux, vêtements ...). Il est ainsi évident que des structures cérébrales qui ne sont pas celles que nous avons décrites plus haut sont capables de garder puis de représenter des signaux visuels, ou du moins de recréer en nous ce qui nous semble exactement semblable à un signal visuel perçu. On mettra aussi dans cette catégorie les images mentales que bâtit un mathématicien qui imagine une configuration géométrique (la suite des entiers naturels ou la topologie d'une surface complexe de Z^4 ou un conducteur qui se remémore le chemin pour se rendre en un lieu passé, un poète qui décrit un jardin ...) Les images hallucinogènes sont intermédiaires entre les figures informes de l'obscurité totale et l'image précise du rêve ou de la conscience, empruntant semble-t-il, d'une part à des aires cérébrales profondes et inconscientes, d'autre part à des structures mentales construites, ordonnées et héritées d'une expérience personnelle. Ces images mentales qui se passent de la vision et qui attribuent l'essentiel de la responsabilité de la perception à la conscience ont fait l'objet de théories avancées, en particulier de Goethe [Goethe, 1792] qui réfutait les interprétations spectrales de la couleur et de Schopenhauer [Schopenhauer, 1966].

Ces phénomènes extrêmes de vision, où l'information optique est faible au regard des antécédents conscients et inconscients expliquent le succès des multiples exemples de "divination" (marc de café, entrailles de poulet, nuages ou fumées, lignes de la main) ainsi que - plus académiquement - des tests de Rorschach. Nous ne nous étendrons pas sur ces aspects qui nous éloignent de la vision par ordinateur, mais nous devons en garder mémoire car elles ilustrent bien le rôle que peut prendre notre cortex dans les manifestations de la vision.

1.2.5 Un modèle algorithmique de vision : le modèle de David Marr

David Marr est un neurophysiologue qui a proposé de décrire la vision par un enchaînement d'opérations purement mécaniques entre des couches de cellules indépendantes. Cette approche, qu'il a étendue à l'ensemble des fonctions de la cognition, a conduit à la naissance des **neurosciences computationnelles** [Marr, 1982] dans les années 1970-1980.

David Marr a été très inspiré par les schémas des colonnes visuelles mises en avant par Hubel et Wiesel [Hubel and Wiesel, 1959]. Il a donc proposé une représentation hiérarchique de la vision qui part de l'image bidimensionnelle présente sur la rétine. Il est également guidé par les travaux conduits sur la machine informatique par Türing. Il en gardera un schéma de traitement en série. Il élabore ainsi un modèle mental tridimensionnel de la scène, réparti sur 3 niveaux :

1. une représentation primaire, dénommée *primal sketch*, fournit une délimitation spatiale des principaux objets de la scène. Elle est le résultat de toutes les étapes de traitement (amplification et détection des contours, détection de zones homogènes, segmentation, séparation chromatique ...) ;
2. une représentation intermédiaire, dénomée *2.5D sketch*, prend en compte des informations volumiques sur les gradients d'éclairage ou de textures (forme à partir de l'ombrage [16]), sur les parties cachées, sur les ombres projetées, mais aussi sur les disparités de mouvement, le flot de mouvement et la stéréovision. Cette représentation, quoique tridimensionnelle puisqu'elle prend en compte la profondeur de la scène, reste centrée sur le repère visuel de l'observateur ;
3. un modèle mental tridimensionnel et continu dans lequel les objets du monde réel prennent place, indépendamment du point de vue de l'observateur, lui permettant de naviguer virtuellement dans la scène et de restituer pour chaque objet une vue centrée sur cet objet et en particulier de rapporter cet objet à sa représentation dans un "catalogue".

La première étape a fait l'objet de très nombreux travaux et s'est montrée assez fructueuse en suivant de très près les schémas de la biologie : traitement hiérarchique de l'information selon les

16. Shape from shading, citedurou :08

orientations fournies par un filtrage fréquentiel pyramidal en ondelettes, amplification, détection contextuelle, etc. Dans le texte original de Marr, elle s'appuie sur une segmentation de l'image à partir des passages par zéro du laplacien filtré par une gaussienne [Burt, 1984]. La gaussienne reflète les filtres obtenus par regroupement des cellules visuelles, tandis que le laplacien de gaussiennes, assez proche d'une différence de gaussiennes, reflète la juxtaposition de cellules OFF. La hiérarchie des échelles des gaussiennes traduit l'empilement des colonnes visuelles.

La première comme la seconde étape se prêtent bien à une approche *bottom-up* qui part des détails fins et construit, par association, des objets composites de plus grandes dimensions. Elles sont bien adaptées à l'architecture des ordinateurs et aux principes de programmation aussi bien séquentielle que parallèle.

La mise en œuvre de la troisième étape est plus délicate et ne semble pas pouvoir être opérationnelle sans avoir recours à des boucles de rétroaction ou des opérations de déduction consciente (de raisonnement) auxquelles souhaitait échapper D. Marr.

Si la communauté du traitement d'images s'est (presque) unanimement engagée dans la voie tracée par D. Marr, elle a progressivement pris ses distances avec une approche purement *bottom-up* à partir des années 2000 et considère aujourd'hui qu'une interprétation de scène efficace ne peut être obtenue qu'en introduisant très tôt une forme de sémantique dans les opérations élémentaires. Cette information de haut-niveau sémantique serait apportée dans le système visuel humain par les boucles de retour des aires supérieures vers le corps genouillé ou les aires visuelles. Les succès des techniques à base de réseaux de neurones préalablement entraînés sur des scènes naturelles dûment étiquetées, ainsi que l'efficacité de certaines achitectures artificielles comportant des rétro-bouclages entre couches, semblent confirmer cette idée.

1.2.6 Une théorie de la perception : la *Gestalttheorie* ou psychologie de la forme

L'approche adoptée par l'école gestaltistes ([Koehler, 1929] ou [Gordon, 1998] chapitre 3) des psychologues expérimentaux des années 1925 est une réponse originale aux recherches entreprises pour définir un cadre formel à la vision. Pour eux, la perception des signaux issus des objets ne s'expliquerait pas par un processus de concaténation ou de sommation en série de leurs différentes parties. Ils seraient perçus comme un tout, sous certaines conditions qu'ils ont cherché à cerner. Cette approche, à l'inverse de la précédente, récuse une démarche *bottom-up* et rend très difficile un traitement parallèle. Ainsi, les 3 segments de la figure 1.12, lorsqu'ils se rapprochent suffisamment jusqu'à se rejoindre, forment une figure triangulaire que l'on perçoit en soit plutôt que comme la somme de trois composantes indépendantes. Cette structure triangulaire n'émerge qu'avec un "recul" suffisant.

Cette construction suivrait un certain nombre de "lois" illustrées sur la figure 1.14. Ces lois reflètent un principe de simplicité qui présiderait aux actions du cerveau afin de limiter au maximum les efforts d'interprétation nécessaires à la reconstruction de l'information. Prenons l'exemple des contours subjectifs de la figure 1.13. Un triangle blanc posé sur le dessin n'a pas de contour réel pourtant on le perçoit très vivement. Il paraît même être plus lumineux que le fond. Or, qu'y a-t-il dans cette image ? Trois angles et trois portions de disques. Ces formes paraissent incomplètes au cerveau. Il suffit alors de faire l'hypothèse supplémentaire de la présence d'un triangle blanc recouvrant et cachant partiellement ces formes pour que l'image s'explique très simplement.

La psychologie de la forme prône donc une interprétation globale de la scène et rejette la possibilité d'une interprétation ascendante (*bottom-up*) à partir de la somme de ses composantes. Malgré sa capacité à bien expliquer l'interprétation que nous faisons de scènes naturelles simples, la théorie de la Gestalt a été laissée de côté lors du développement du traitement numérique de l'image et de la vision artificielle (en particulier parcequ'elle est incompatible avec la théorie de D. Marr). Mais la Gestalt a retrouvé une place importante en vision par ordinateur en apportant des fondements très fructueux à des opérations difficiles comme la détection de contours dans les signaux très bruités,

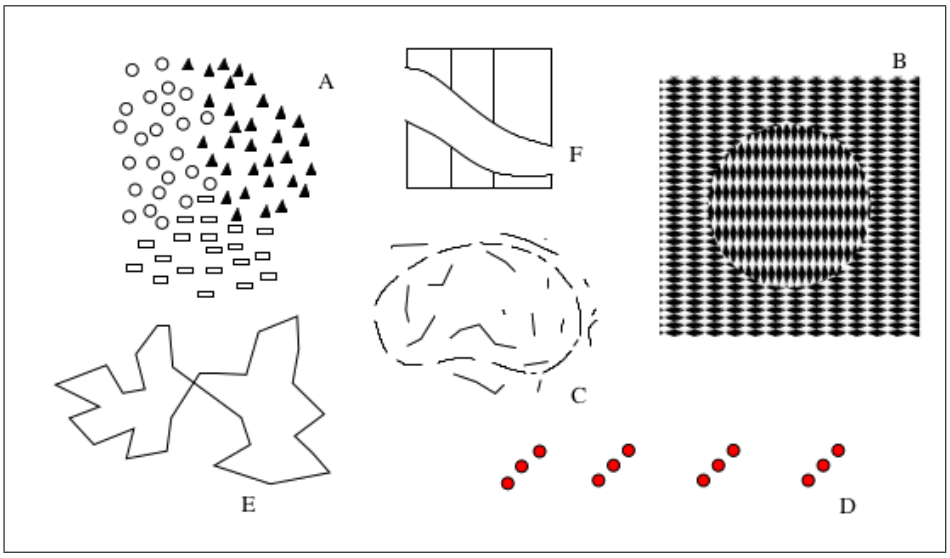

Figure 1.14 – *Lois de la perception (d'après les gestaltistes). En A, loi de similitude : les objets élémentaires ayant des caractéristiques similaires sont regroupés en trois zones. En B, loi d'orientation : les deux textures sont séparées et un objet circulaire est créé qui n'a pas de contour propre. En C, loi de continuité : les morceaux de contours qui permettent de former un contour fermé sont spontanément reliés et l'on fait abstraction des autres. En D, loi de proximité : l'interprétation la plus fréquente est "4 triplets de 3 points" plutôt que "3 lignes de 4 points". En E, loi de fermeture : on perçoit deux formes fermées se touchant en un point plutôt que deux courbes quelconques se croisant. En F, loi de continuité : au lieu de percevoir 2 formes côte à côte, on préfère voir une forme courbe recouvrant trois formes rectangulaires juxtaposées.*

la recherche de signaux faibles ou la reconnaissance de formes incertaines (par exemple en imagerie médicale). Ces développements adoptent des approches probabilistes pour estimer la possibilité de rencontrer des configurations particulières en raison soit de l'aléatoire du bruit dans l'image, soit de l'existence d'une "forme" sous-jacente [Desolneux et al., 2008].

L'approche gestaltiste permet également de bien prendre en compte les particularités physiologiques de la circuiterie des aires visuelles, tout en laissant la place à des facteurs inconscients qui sont issus du cortex non-visuel (mémoire, culture). A ce titre elle est plus compatible avec les résultats récents des "sciences de l'esprit".

1.3 Les illusions d'optique

Le système visuel offre des performances remarquables, mais, comme nous l'avons signalé, ce n'est pas un instrument de mesure passif du flux visuel. Il peut transmettre aux aires visuelles des signaux différents de ceux que l'on attendrait d'un système d'imagerie passif comme un appareil photo. Nous avons vu le rôle de pré-filtrage des voies optiques, nous avons donné des explications en termes de formes, mais il faut aussi faire intervenir des couches plus lointaines de notre conscience, issues de notre expérience, de notre éducation ou de notre culture qui interviennent dans notre façon de "lire" une scène. Ces traitements complexes se révèlent en particulier à travers ce que l'on nomme de façon très générique les **illusions d'optiques**, mais qui ne sont pas toutes des illusions.

Ces différences peuvent provenir de divers niveaux du traitement des signaux visuels :

— les bandes de Mach (figure 1.10, les illusions de Titchener, de Muller-Lyer, de Jastrow (figure 1.15), l'illusion de Zollner, les spirales de Searl et, dans une certaine mesure, l'échiquier d'Adelson (figure 1.15), mettent en jeu les toute-premières couches de la rétine et montrent l'influence du contexte immédiat soit dans l'estimation des longueurs, soit dans l'appréciation des nuances et des couleurs. Elles montrent également l'hétérogénéité de la rétine et de ses

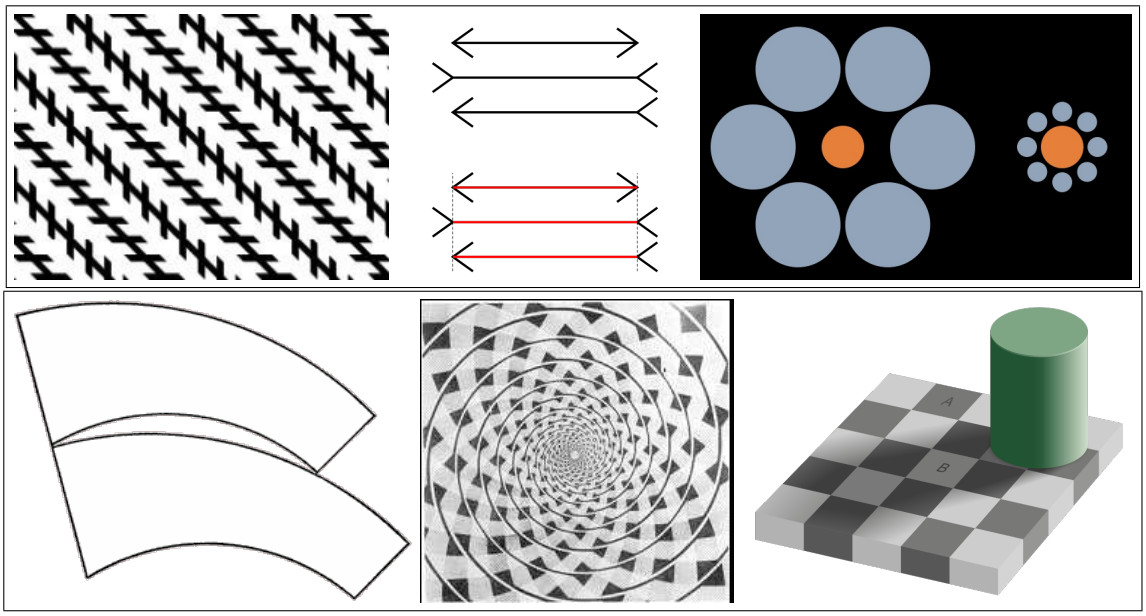

Figure 1.15 – *Phénomènes dûs aux interactions spatiales dans l'image. En haut, de gauche à droite : les diagonales de Zollner sont parallèles contrairement aux apparences ; les flèches de Muller-Lyer sont de même taille, mais celles du centre apparaissent plus longues ; dans l'illusion de Titchener, les deux cercles au cœur des fleurs ont même taille. En bas, de gauche à droite : dans l'illusion de Jastrow : les deux bordures (celle supérieure de la forme supérieure et celle inférieure de la forme inférieure) la spirale de Snearl est constituée de cercles concentriques ; dans l'illusion de Jastrow : les deux bordures (celle supérieure de la forme supérieure et celle inférieure de la forme inférieure) sont de tailles identiques ; dans l'échiquier d'Adelson, les deux carrés marqués A et B ont le même niveau de gris, pourtant B apparaît blanc et A noir (©E. H. Adelson).*

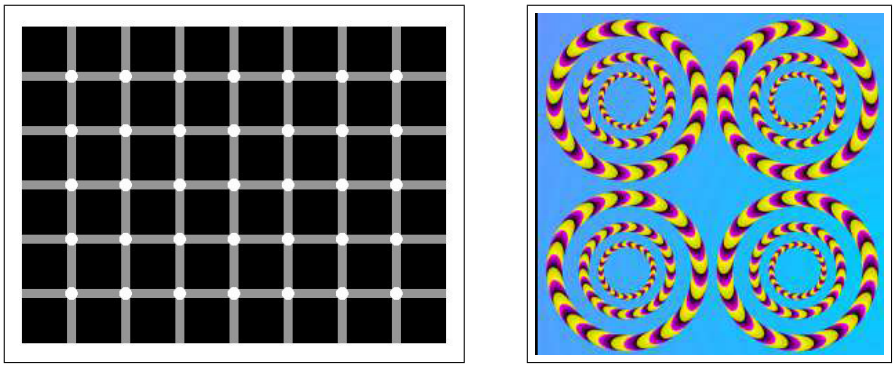

Figure 1.16 – *Dans ces illusions entrent en jeu les différences de traitement (en couleur et en résolution) entre la fovéa et les zones périphériques de l'œil : à gauche, dans les grilles d'Hermann, des taches noires apparaissent aux croisements des lignes blanches pour les seuls régions où ne se fixe pas l'attention de l'observateur. Ces taches changent de position quand on change de point de fixation (phénomène dit* de mouches*). à droite, Le mouvement apparaît sur les cercles que l'on ne fixe pas, en raison du très fort contraste coloré qui est moyenné en périphérie de l'œil et pour lequel les moyennes sont prises sur des voisinages différents lorsque la fixation change.*

populations de cellules réceptrices. Ces illusions nous conduiraient à une décision erronée sur une mesure (longueur, taille, niveau de gris).

— la grille d'Hermann ou les cercles de Fraser (figure 1.16), utilisent des mécanismes intermédiaires, où sont confrontées les propriétés de la vision fovéale et de la vision périphérique. Ils font apparaître des signaux inexistants (taches noires aux carrefours) ou des mouvements absents. Ils sont très sensibles aux mouvements de la tête et au déplacement du regard.

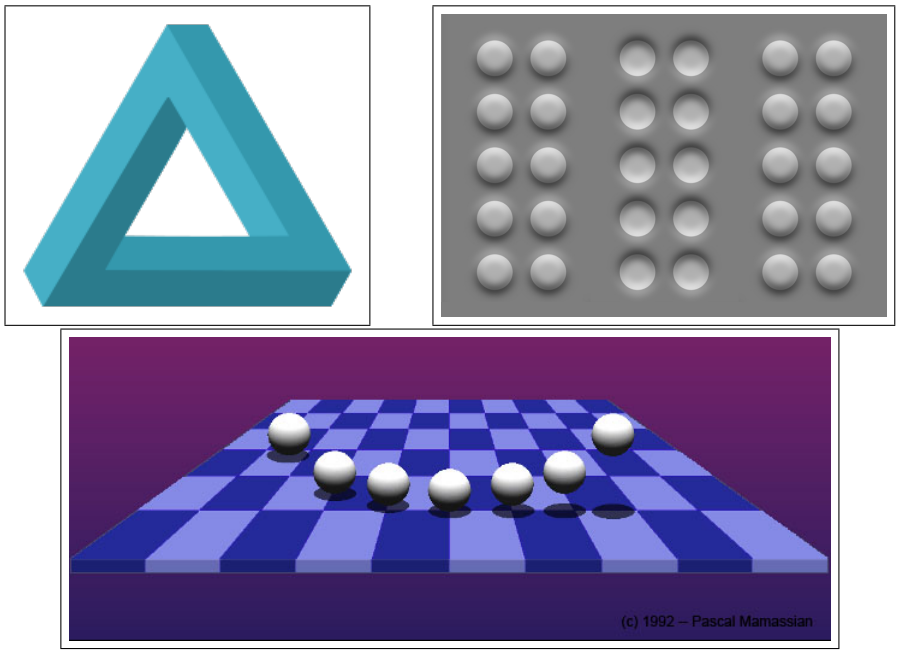

Figure 1.17 – *Les phénomènes illustrés ici ne sont pas à proprement parler des illusions d'optique. Ils présentent l'importance des effets cognitifs et culturels dans l'interprétation des images. En haut, à gauche, le triangle de Penrose est la figure visuellement possible d'un objet physiquement impossible de l'espace tridimensionnel. L'observateur cherche une représentation tridimensionnelle qu'il ne trouve pas sans se contenter de l'image bidimensionnelle. Il y a conflit entre le message transmis et ce que notre être conscient souhaite voir ; En haut à droite, des variations d'éclairage conduisent à des interprétations subjectives de relief. On interprète généralement les cercles à gauche et à droite comme des déformations de la surface vers l'observateur tandis que ceux du centre sont interprétés comme des creux. Les raisons de ces interprétations sont culturelles. Nous sommes habitués à l'éclairage solaire venant du haut et notre cerveau propose spontanément de lire les images sous cette hypothèse qui nous reste inconsciente ; on "voit" donc différemment des images qui sont identiques mais symétriques ; En bas, dans le billard de Mamassian, les boules sont dessinées symétriquement à droite et à gauche, mais un observateur associe une tache à chaque boule qu'il interprête comme son ombre, et l'image est alors "vue" de façon disymétrique : les boules à droite s'élèvent au dessus du billard tandis que celles à gauche reculent vers le fond du billard.*

— le triangle de Kanisza (figure 1.13) mêle les propriétés des colonnes visuelles à une analyse globale gestaltiste, une interprétation spontanée est faite au titre de l'interprétation. Elle entraîne en retour une modification de la perception de bas niveau en lui faisant "voir" un triangle blanc au dessus des autres tracés.
— Le triangle de Penrose (figure 1.17), l'éléphant à cinq pattes, les cascades infinies d'Escher ... nous proposent des dessins que l'on voit sans erreur mais dont nous ne pouvons faire une interprétation mentale car ils ne peuvent correspondre à aucun objet possible dans le monde réel. Ce sont des objets impossibles auxquels on oppose des mécanismes de vérification de cohérence de haut niveau. Ces mécanismes sont alors en échec sans qu'il soit possible d'impliquer les traitements antérieurs. Nous confrontons un signal perçu à un dictionnaire de formes constitué par notre expérience passée et cette confrontation échoue.
— Le champ de bosses (figures 1.17), le cube de Necker, la vieille/jeune-femme (figure 1.18), le canard/lapin, proposent des images pouvent donner lieu à deux interprétations mentales différentes (le cube est vu par dessus ou par dessous). Dans cette situation, les deux interprétations ne sont généralement pas également probables. L'une s'impose souvent pour des raisons culturelles (c'est le cas du champ de cratères). Il est à noter que l'on passe alors de l'une à l'autre des interprétations par un effort volontaire de conscience. Il faut également noter que, même lorsque les deux interprétations sont également probables - par exemple pour le cube de Necker - il n'est pas possible d'en disposer simultanément. A un instant donné, on

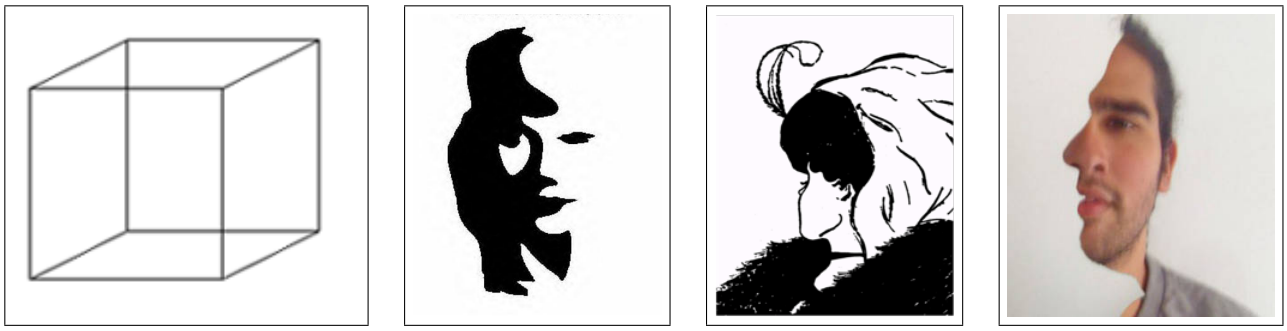

Figure 1.18 – *Images ambiguës. Ces images présentent deux interprétations duales qui correspondent exactement au même signal optique, qui subit les mêmes prétraitements par les voies visuelles, mais qui correspondent à deux compréhensions différentes. Chaque interprétation est très stable (elle n'est pas affectées par de petites modifications de l'image ou des variations de l'attention). Mais ces interprétations sont incompatibles. On ne peut percevoir l'une et l'autre simultanément et l'on ne passe de l'une à l'autre qu'à la suite d'un effort mental et certaines personnes peuvent ne pas percevoir certaines interprétations. Chaque interprétation est donc probablement le résultat d'une longue suite de traitements et l'on ne repasse de l'une à l'autre que par un long retour en arrière dans le traitement visuel des indices qui conduisent au sens.*

voit soit l'une soit l'autre.

— La constance chromatique qui nous rend très peu sensibles à l'influence chromatique de l'illuminant dans la reconnaissance d'un objet coloré (on "voit" les couleurs d'un objet par rapport aux couleurs de ceux qui l'entourent) et qui conduira à la *balance des blancs* en photographie, peut être considérée comme une illusion d'optique. C'est également un élément qui participe à l'apparence des gris dans l'échiquier d'Adelson (figure 1.15).

1.4 La vision stéréoscopique

En 1838 Wheatstone a montré que la stéréovision permet une reconstruction tridimensionnelle à partir des deux images acquises par les deux yeux. Vers 1965, des expériences ont mis en évidence des cellules des voies optiques répondant spécifiquement à la disparité spatiale, montrant que cette capacité est en partie biologiquement cablée.

La tête étant verticale, lorsque le regard est à l'infini, les lignes épipolaires [17] sont horizontales. En vision à distance finie, ces lignes épipolaires forment des faisceaux de droites convergentes au point de fixation. Les positions des images i_d et i_g d'un même objet I sur leur droite épipolaire diffèrent de la disparité absolue ($\Delta_I = i_d - i_g$) qui est fonction de la base stéréoscopique et de la distance de I à l'observateur. En raison du mouvement constant du regard, cette disparité absolue varie constamment. La disparité relative entre deux objets $\delta(i,j) = \Delta_J - \Delta_I$ est cependant une constante pour une position donnée de la tête et une fixation à distance donnée. C'est elle qui porte l'information nécessaire à la reconstruction stéréo du point I par rapport au point de fixation [Neri, 2005].

Après le chiasma optique, les deux images droite et gauche de I sont réunies [18] dans un même faisceau de fibres et contribuent conjointement à une même région de la carte rétinotopique. On trouve alors, dès les aires primaires du cortex visuel, des cellules qui réagissent en fonction de la disparité et qui collectivement couvrent la totalité des dynamique de disparité selon trois modes : celles qui agissent (ou s'inhibent) en présence d'une disparité nulle, d'une disparité positive ou d'une disparité négative [Poggio, 1995]. Ces cellules plongent leurs champs récepteurs autour du même point de

17. En vision stéréoscopique, les deux images I' et I" d'un point I de l'espace à 3D sur les plans images, forment un plan de l'espace 3D qui coupe les 2 plans images selon des lignes dites épipolaires. La distance du point I à l'observateur peut être déduite des positions relatives de I' et I" sur ces droites épipolaires. La géométrie épipolaire (qui rapporte tous les points des deux images à leurs lignes épipolaires), permet de réduire la recherche du correspondant d'un point, à une recherche à 1D et non à une recherche à 2D beaucoup plus coûteuse [Faugeras, 1993].

18. par moitié à droite et à gauche.

disparité nulle dans chacun des deux champs visuels. Leur réaction semble fondée sur une similarité texturale assez semblable à une corrélation [Goutcher and Hibbard, 2014]. Ces mesures locales de disparité sont transmises aux aires ultérieures du cerveau. Quoique l'on trouve dès l'aire V1 puis dans les aires suivantes des cellules aptes à exploiter la disparité stéréoscopique, des travaux actuels semblent attribuer à une aire assez tardive dans le traitement visuel, l'aire A4 en charge, entre autres, du contrôle des mouvements de l'œil, l'essentiel de la fonction de stéréovision [Neri, 2005]

Les études sur la stéréovision ont grandement bénéficié de l'invention des stéréogrammes aléatoires [Julesz, 1971] (figure 1.19). Ils ont permis à B. Julesz d'étudier systématiquement les capacités de discrimination 3D de l'observateur humain dans des expériences qui ne font pas appel à une compréhension sémantique de la scène. Ces expériences ont permis d'écarter l'explication de la stéréovision par des appariements de haut niveau après interprétation de la scène.

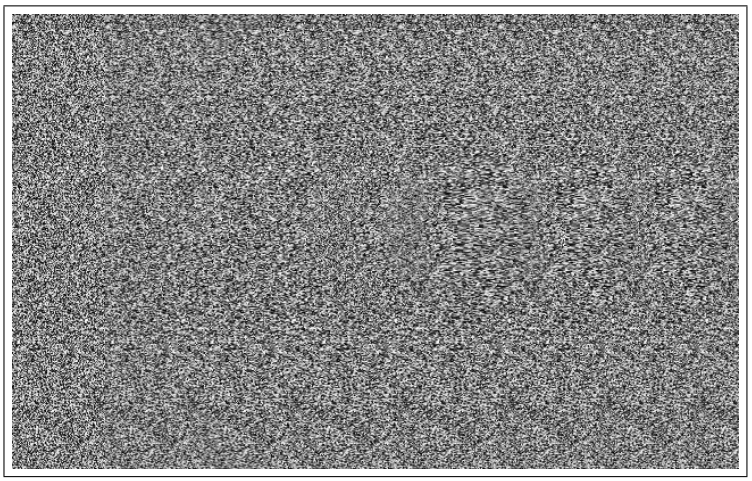

Figure 1.19 – *Auto-stéréogramme aléatoire (*random-dot self-stereogram*). Un stéréogramme aléatoire présente deux images séparées, l'une vue par l'œil droit, l'autre par l'œil gauche. La fusion des deux images fait apparaître un relief recouvert d'une texture aléatoire. Un auto-stéréogramme (comme celui-ci) regroupe les deux images en une seule. La lecture d'un auto-stéréogramme nécessite de désolidariser les fonctions de mise au point et de convergence (une opération qui n'est pas naturelle et nécessite quelqu'entraînement). Pour ce stéréogramme observé à la distance de sa diagonale environ, il faut que les deux axes de vision soient parallèles tandis que la mise au point se fait sur la texture. Généralement trois images apparaissent dans le champ de vision. Celle du centre est l'image fusionnée en relief. Elle représente ici un "volcan" vu de dessus. ©jchr.be*

1.5 La couleur : modélisation et propriétés

Nous avons très brièvement évoqué plus haut la découverte par Grassman, en 1853, de la trivariance de la couleur pour l'observateur normal et l'additivité des couleurs. Ses expériences ont montré que toute couleur du spectre visible (du violet autour de 400 nm jusqu'au rouge profond vers 750 nm) pouvait être reconstruite à l'aide de trois longueurs d'ondes pures correctement mélangées. Pour cela, Grassman a projeté sur un écran la couleur inconnue et a juxtaposé à cette couleur la superposition de trois plages issues des primaires. En réglant des atténuateurs sur les trois primaires, il obtient une couleur identique à la plage inconnue. Toute couleur est donc bien perçue de façon équivalente à la somme pondérée des trois primaires, les poids se déduisant des atténuations appliquées [19]. Une autre découverte très importante a été faite en 1886 par Abney : la linéarité en luminosité des processus additifs de combinaison des couleurs : si A a même luminosité que X, et B même luminosité que Y, alors A+B a même luminosité que X+Y.

19. En fait, comme nous le verrons plus bas, la manip s'avère un peu plus délicate et l'équilibre n'est pas toujours possible.

Enfin, la possibilité de créer une couleur soit par un processus additif (partant du noir et ajoutant des "lumières" comme dans l'expérience de Grassman), soit par un processus soustractif (partant du blanc en supprimant des composantes chromatiques, comme dans les films) a été mise en évidence.

Le besoin de préciser, décrire, reproduire, transmettre, comparer des couleurs précises a amené la communauté scientifique à fixer le choix de l'espace tridimensionnel le plus adéquat pour les divers besoins et d'en étudier les propriétés.

De telles études auraient dû s'appuyer sur une connaissance précise des capteurs à l'œuvre dans la vision humaine (les courbes LMS que nous avons vues en section 1.5) mais ces connaissances n'étaient pas disponibles en 1920 quand ces travaux ont commencé [20]. La colorimétrie universellement utilisée a été bâtie dans les années 1930 et régulièrement remise à jour en fonction des progrès des connaissances et des besoins nouveaux des industries (la photo couleur, la télévision à tube, les capteurs solides, les écrans plats ...). Elle est construite expérimentalement à partir d'un toute petit nombre d'observateurs (7 initialement) qui ont reproduit les expérimentations de Grassman dans des conditions très contrôlées (en particulier en ce qui concerne l'éclairage d'ambiance, les niveaux lumineux, l'angle solide des plages de contrôle, la durée d'observation). La colorimétrie a été normalisée par un ensemble de formalismes établis sous l'égide de la CIE (Commission Internationale de l'Eclairage) de façon à pouvoir reproduire identiquement des expériences perceptuelles pour des industries variées. Les expériences fondatrices de 1931 ont été répétées et améliorées après guerre pour s'adapter au monde de la photographie et de la télévision en couleur. L'observateur CIE 1964 a remplacé observateur CIE 1931 dont il diffère cependant assez peu. Les anomalies de la vision colorée (daltonisme) ont été également finement étudiées.

1.5.1 L'espace RVB de la CIE 1931

La CIE a donc précisé les expériences permettant de définir les primaires des expériences de Grassman. Elle a ainsi défini un espace RVB de référence (CIE-RVB 1931) [21] à partir des éléments suivants :

- les trois primaires sont monochromatiques, de longueurs d'onde :
  - 700 nm pour le rouge,
  - 546 nm pour le vert,
  - 436 nm pour le bleu ;

  Des fréquences pures ont été choisies de façon à restituer un vaste ensemble de couleurs par des sommations positives. Leur position a été déterminée de façon à bien couvrir les domaines de forte sensibilité du système visuel ;
- un blanc de référence est choisi, dénoté E ou W_0, de densité de puissance constante en longueur d'onde et égale à 5,3 10^{-2} W/nm ; c'est le blanc d'égale énergie.

Un observateur idéal (moyenne des 7 expérimentateurs) a été défini à partir de ces éléments, qui se caractérise par ses fonctions colorimétriques (ou chromatiques), exprimant la façon dont toute longueur d'onde se décompose en trois composantes.

Comme dans les expériences de Grassman, chaque observateur procède par égalisation de plages chromatiques. Il s'efforce d'égaliser une longueur d'onde pure qui va balayer tout le spectre des longueurs d'onde visibles, à l'aide d'un mélange des 3 primaires. Si l'équilibre est obtenu, on écrit :

une longueur d'onde pure = mélange pondéré des trois primaires

Si l'équilibre ne peut être obtenu, il ajoute l'une des primaires à la longueur d'onde inconnue de façon à obtenir l'équilibre et l'on convient de compter négativement la quantité de cette primaire :

20. Une colorimétrie LMS se développe depuis quelques années, mais, arrivant sur un terrain déjà largement occupé par les standards anciens, en particulier en milieu industriel, elle est encore peu utilisée [CIE, 2006, Vienot and Le Rohellec, 2012] en raison de la popularité des colorimétries fondées sur les observateurs CIE.

21. Il sera corrigé légèrement par la CIE en 1964 (observateur standard CIE 1964)

une longueur d'onde pure + une primaire = mélange approprié des deux autres primaires

soit :

une longueur d'onde pure = mélange de deux primaires - une portion de la troisième primaire.

Les courbes résultantes montrant la portion de chaque primaire pour toute longueur d'onde pure sont présentées sur la figure 1.20, elles sont dénotées : $\overline{r}(\lambda)$, $\overline{v}(\lambda)$, $\overline{b}(\lambda)$. En fait, elles présentent presque partout [22] au moins une composante négative, mais ces composantes sont très faibles hormis dans la région du vert où elles se traduisent par une forte contribution négative du rouge. Comme on le constate, ces courbes sont assez notablement différentes de celles traduisant la sensibilité en longueur d'onde des capteurs l'œil humain (figure 1.3) qui, elles, sont toujours positives.

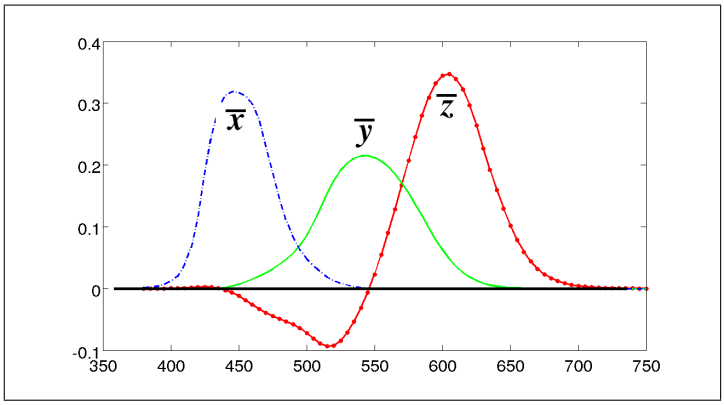

Figure 1.20 – *Fonctions colorimétriques de l'observateur CIE de référence [CIE, 2006]. Ces courbes sont assez différentes de celles obtenues en isolant chacun des pigments notés LMS de l'œil (figure 1.3) : elles couvrent plus uniformément l'intervalle chromatique.*

Mais la plupart des sources de lumière ne sont pas des longueurs d'onde pure. Si l'on considère un stimulus quelconque décrit par son spectre en fréquence $\Phi(\lambda)$, ses composantes trichromatiques R, V et B sont données par :

$$\begin{array}{l} R = A \int \Phi(\lambda)\overline{r}(\lambda)d\lambda \\ V = A \int \Phi(\lambda)\overline{v}(\lambda)d\lambda \\ B = A \int \Phi(\lambda)\overline{b}(\lambda)d\lambda \end{array} \tag{1.3}$$

où l'intégrale est prise sur tout le domaine visible, et la valeur de normalisation A vaut 1 si $\Phi(\lambda)$ est exprimé en watt et 683 s'il est exprimé en lumen (683 lumen/watt correspond au maximum de la sensibilité de l'œil). On peut également utiliser les coordonnées normalisées par division par la somme $R + V + B$, on les écrit alors en minuscules :

$$\begin{array}{l} r = \frac{R}{R+V+B} \\ v = \frac{V}{R+V+B} \\ b = \frac{B}{R+V+B} \end{array} \tag{1.4}$$

C'est le prototype de toutes les représentations à partir de primaires réelles.

L'examen de cette représentation conduit aux conclusions suivantes. Dans le repère RVB, les couleurs accessibles en combinant de façon positive les trois primaires appartiennent à un cube (figure 1.21 à gauche). C'est l'espace des couleurs accessibles avec ces 3 primaires.

22. Cela est dû à la convexité du lieu des fréquences pures qui est toujours extérieur au triangle des 3 primaires, sauf aux 3 primaires elles-mêmes (figure 1.22).

Les tons neutres, de luminosité croissante, s'étalent sur la première diagonale, du noir (0,0,0) au blanc (1,1,1).

Le triangle de Maxwell est le triangle diagonal défini, dans ce cube, par l'équation $R+V+B=1$. Les longueurs d'onde pures forment une surface dans l'espace à 3D. Cette surface intercepte le plan du triangle de Maxwell selon une courbe joignant le rouge profond au bleu/violet, en passant par le bleu, le vert et le jaune. Cette courbe est ouverte dans son chemin direct entre le rouge et le bleu (il n'y a pas de longueur d'onde pure correspondant aux pourpres, mélanges de rouge et de bleu).

Dans tout plan diagonal [23], d'équation $R+V+B=K$ il y a une courbe des longueurs d'onde pures d'intensité indexée par K. Nous l'avons vu, cette courbe pour K variant de 0 à 1 décrit une surface extérieure au cube (sauf le long des trois axes, pour les couleurs correspondant aux trois primaires).

Dans un plan diagonal (par exemple pour $K=1$), il suffit de deux variables pour représenter toute couleur. On construit ainsi un diagramme chromatique qui permet de faire abstraction de la luminosité de l'image. On choisit traditionnellement pour axes R et V (figure 1.21 à droite). Sur ce diagramme, les couleurs accessibles avec les primaires R, V et B sont toujours situées dans le triangle (0 :1, 0 :1). Le lieu des longueurs d'onde pures est extérieur à cette portion de l'espace.

Simple de conception, cet espace CIE-RVB est avant tout un espace de référence, mais il est néanmoins très peu utilisé car il est mal-commode d'emploi car il comporte des composantes négatives. Pour cela, on lui préfère celui construit à partir de l'espace XYZ (voir figure 1.22).

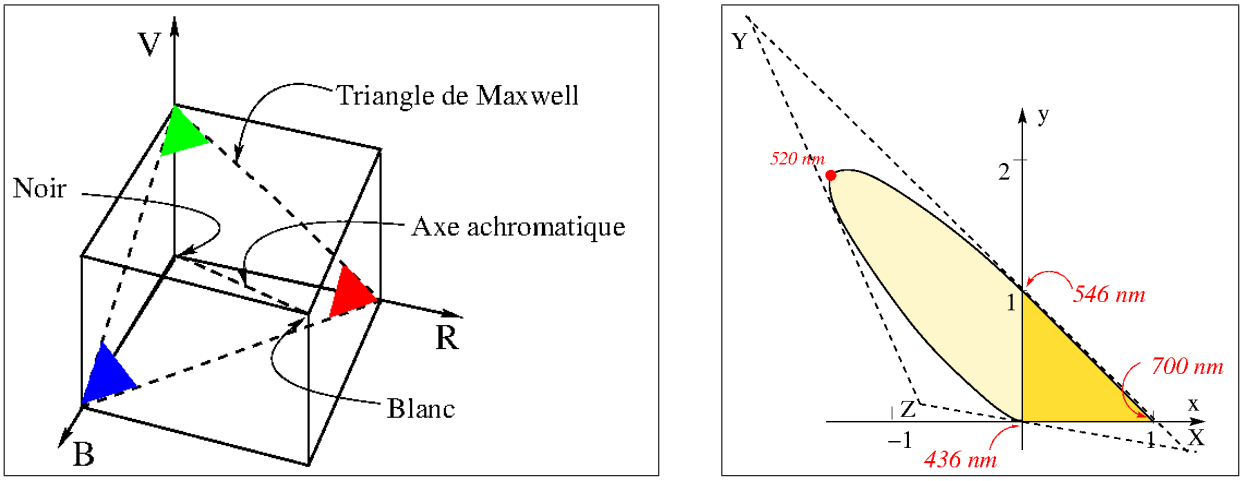

Figure 1.21 – *A gauche : le cube des couleurs RVB et le triangle de Maxwell. A droite le diagrame chromatique (x,y) construit sur RVB à luminosité constante (R+V+B = K) : on place l'origine au point B de la figure de gauche et on conserve les seules coordonnées R et V normalisées. Le lieu des fréquences pures comporte une très importante partie d'abscisse négative. On préfère changer d'espace de façon à n'avoir que des coordonnées positives. L'espace CIE-RVB reste Le triangle XYZ est celui qui a été retenu pour définir les axes de l'observateur CIE 1931 à cette fin.*

1.5.2 La représentation XYZ

On a donc élaboré un autre espace lui aussi artificiel, mais plus commode d'usage, par un changement de repère.

Un axe a été choisi de façon à exprimer la luminance de l'image. Puis les deux autres axes ont été pris de façon que toute décomposition d'un stimulus coloré soit positive. Pour cela on a choisi pour nouveaux axes des tangentes particulières au lieu des couleurs pures (voir la figure 1.21 à droite). Le lieu des fréquences pures étant convexe, un tel choix garantit cette positivité.

23. Un tel plan est souvent dit à énergie constante ou à luminosité constante quoique ces termes doivent être pris avec précaution.

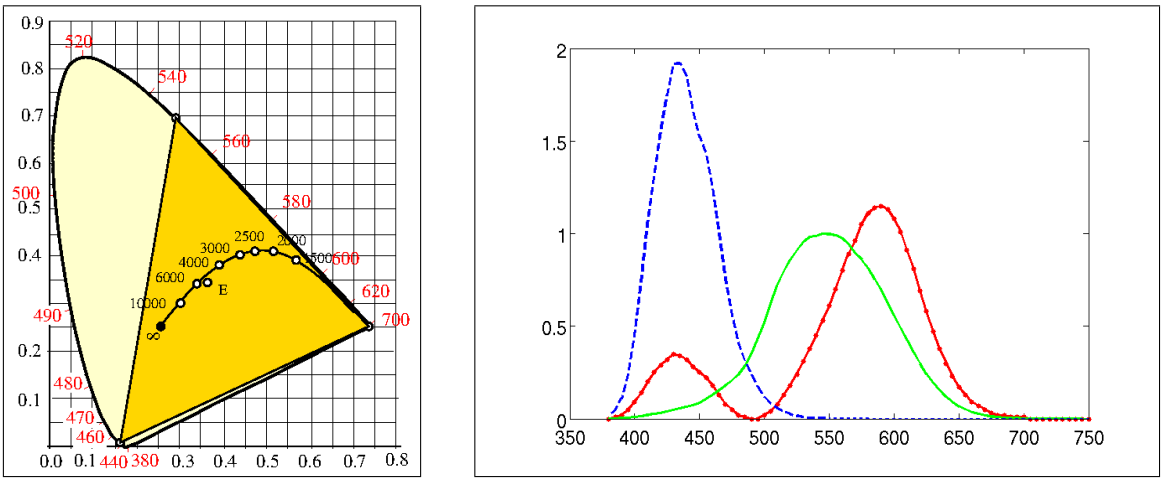

Figure 1.22 – *A gauche, le diagramme de chromaticité de l'observateur CIE 1931. Il se distingue de celui de la figure 1.21 car toutes les couleurs ont ici des coordonnées positives. La courbe extérieure au triangle est le lieu des fréquences pures, allant ici de 380 nm (bleu profond) à 700 nm (rouge) ; elle est limitée dans sa partie basse par la droite des pourpres. Le triangle est celui des couleurs atteignables par des sommes positives des trois primaires RVB choisies par la CIE. On explique par ce diagramme que la composante négative des trois primaires $\overline{x}$, $\overline{y}$, et $\overline{z}$ n'est forte que pour le domaine vert-jaune (voir figure 1.20), mais théoriquement bien présente pour toute longueur d'onde pure. Le lieu des corps noirs est représenté par la ligne traversant le domaine, graduée en températures (Kelvin). Le blanc équiénergétique (en longueur d'onde) W_0 est noté E. A droite, les fonctions colorimétriques associées à cet espace XYZ. Notons qu'elles sont toujours positives. Elles sont très éloignées des courbes d'absorption des photorécépteurs LMS (figure 1.3)*

Il est ainsi possible de définir un diagramme de chromaticité à partir de l'espace XYZ (figure 1.22). Dans ce diagramme, on choisit pour variables x et y à partir des deux dimensions X et Z de l'espace XYZ. C'est la représentation universellement adoptée pour représenter un point de couleur, malgré quelques défauts résiduels que nous verrons plus loin.

Utilisant les fonctions colorimétriques $\overline{x}(\lambda)$, $\overline{y}(\lambda)$ et $\overline{z}(\lambda)$ de la figure 1.20 :

$$\begin{array}{l} X = A \int_{visible} \Phi(\lambda)\overline{x}(\lambda)d\lambda \\ Y = A \int_{visible} \Phi(\lambda)\overline{y}(\lambda)d\lambda \\ Z = A \int_{visible} \Phi(\lambda)\overline{z}(\lambda)d\lambda \end{array} \tag{1.5}$$

où, comme pour l'équation (1.4), A vaut 1 ou 683 selon que le flux est exprimé en watt ou en lumen, on passe de RVB à XYZ par un changement de base :

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0,4887 & 0,3107 & 0,2006 \\ 0,17620 & 0,81298 & 0,010811 \\ 0,00 & 0,01020 & 0,98979 \end{bmatrix} \begin{bmatrix} R \\ V \\ B \end{bmatrix} \tag{1.6}$$

et inversement de XYZ à RVB par :

$$\begin{bmatrix} R \\ V \\ B \end{bmatrix} = \begin{bmatrix} 2,37067 & -0,9000 & -0,47063 \\ -0,51388 & 1,42530 & 0,08858 \\ 0,055298 & -0,14695 & 1,00939 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \tag{1.7}$$

La représentation matricielle des équations 1.6 permet très aisément de changer de primaires et de nombreux espaces de représentation dérivés de CIE-1931 ont ainsi été proposés.

L'espace sRVB - L'espace RVB a été créé à partir de 3 fréquences pures. Ce choix est excellent pour couvrir au mieux le domaine de Maxwell. Mais d'un point de vue pratique, ces primaires ne sont pas aisées à manipuler car, obtenues par des raies d'émission particulières, elles ont peu d'énergie.

Les industriels ont proposé d'utiliser des primaires en dehors du lieu des fréquences pures (on perdra alors la capacité d'engendrer des couleurs très saturées) mais plus faciles d'emploi pour des ingénieurs. C'est ainsi qu'est né le système sRVB (standard RVB) très adapté pour caractériser la couleur des écrans cathodiques. sRVB (ou sRGB en anglais) utilise des primaires un peu désaturées (représentées sur les figures 1.24) :

— Rouge : $X_r = 0,64$ - $Y_r = 0,33$
— Vert : $X_v = 0,30$ - $Y_v = 0,60$
— Bleu : $X_b = 0,15$ - $Y_b = 0,06$

et un blanc sur le lieu du corps noir, á 6 500K (dénommé D65) : $X_w = 0,3127$ - $Y_w = 0,3290$. Pour d'autres applications (par exemple pour l'imprimerie) on préfère d'autres standards comme AdobeRVB, qui couvre un domaine plus large. Les domaines couverts par ces primaires sont décrits sur la figure 1.23. Avec le développement de la photo numérique, en particulier de synthèse, des espaces encore plus vastes ont été proposés à partir de primaires extérieures au lieu des fréquences pures (comme ProPhoto RGB). Bien sûr, le passage entre ces espaces se fait par des matriçages.

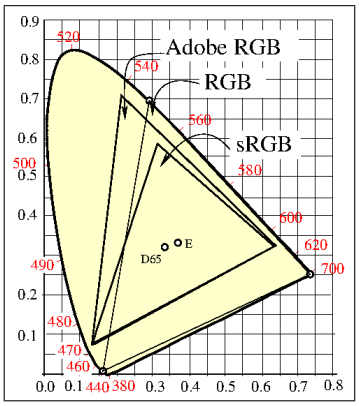

Figure 1.23 – *Afin de faciliter la diffusion très large de normes professionnelles, on préfère utiliser des primaires un peu moins saturées que RVB mais permettant d'avoir plus aisément des sources puissantes. C'est ainsi qu'ont été choisis les primaires de sRVB et d'AdobeRVB.*

Par ailleurs, on peut passer de l'espace LMS à l'espace XYZ par le changement de variables :

$$\begin{bmatrix} L \\ M \\ S \end{bmatrix} = \begin{bmatrix} 0,15514 & 0,54312 & -0,03286 \\ -0,15514 & 0,45684 & 0,03286 \\ 0 & 0 & 0,01608 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \tag{1.8}$$

C'est actuellement la piste qui est utilisée pour redéfinir une colorimétrie qui s'affranchirait de conventions de primaires.

1.5.3 Quelques propriétés des espaces trichromatiques

Avant de quitter l'espace RVB, nous allons examiner certaines propriétés qu'il partage avec de nombreux autres espaces de couleur et que nous rencontrerons donc par ailleurs.

Nous introduisons ici quelques termes qui nous seront utiles et qui reflètent des propriétés importantes des espaces trichromatiques.

Métamérisme

La trivariance de la vision humaine explique l'existence du métamérisme. Deux couleurs sont métamères lorsqu'elles sont représentées par le même point de couleur mais sont physiquement constituées de répartitions spectrales différentes. Le métamérisme traduit une transformation irréversible de la vision puisqu'il n'est plus possible de distinguer ces deux points représentés par le même triplet.

C'est cependant un atout considérable puisque cela permet de reconstruire une apparence donnée sans connaissance de la composition spectrale sous-jacente, par un simple ajustement des trois primaires (figure 1.24 à gauche) [24].

Gamut

Les primaires étant fixées, il est possible de déterminer tout l'espace chromatique balayé par un système de reproduction d'image, en faisant varier chaque primaire de zéro à son énergie maximale. Un tel volume s'appelle le gamut, c'est-à-dire la gamme des couleurs représentables à l'aide des primaires. Une question délicate apparaît lorsqu'il faut représenter des couleurs qui sont situées hors du gamut. Si le point de couleur à reproduire se trouve à l'extérieur du gamut, il sera souvent rabattu (c'est-à-dire projeté sur la frontière du volume), donc représenté par une teinte proche moins saturée, à moins que l'on préfère modifier l'ensemble des couleurs en réduisant le contraste et la saturation, de façon à ramener tous les points de couleur dans l'espace accessible (figure 1.25 à gauche).

Mappage tonal

L'opération qui consiste à faire correspondre un ensemble de couleurs au gamut d'un appareil (écran, imprimante) s'appelle le mappage tonal (*ton mapping* en anglais). Il prend une grande importance en particulier pour les travaux sur la haute dynamique de rendu (HDR). Il fait l'objet d'études très intenses dans les industries de la photo-impression.

Lieu des corps noirs

Les corps noirs émettent un rayonnement fonction de leur température selon la loi de Planck. A chaque température correspond un spectre et un point de couleur. Le lieu des corps noirs [25] traverse l'espace des couleurs depuis le rouge le plus profond jusqu'à un point bleu approximativement en $(x = 0,24, y = 0,23)$ (figure 1.22). Il peut être gradué en températures et on y trouve les blancs de référence notés Dxx, xx exprimant la température en centaines de Kelvin (par exemple D65 à 6 500 K) Le point équi-énergie E, fréquemment utilisé comme référence de blanc, se trouve également près de ce lieu et de la température 5 600 K (figure 1.22 à gauche). Comme plusieurs blancs de référence sont situés près du lieu du corps noir mais pas exactement sur lui, on définit des courbes d'iso-températures de couleur de part et d'autre de ce lieu. Notons que ces courbes, tracées dans le diagramme CIE 1931, ne sont pas orthogonales au lieu des corps noirs.

1.5.4 Limites de l'espace RVB

L'espace RVB (ou l'espace XYZ qui en est déduit) n'est pas le meilleur espace pour reproduire la perception visuelle humaine car :

- il est difficile d'attacher simplement une couleur à un point de cet espace ;
- les trois composantes RVB sont perceptuellement fortement corrélées (diminuer la composante verte fait apparaître la teinte plus rouge) ;
- il est difficile de séparer les notions d'intensité et de chromaticité ;
- la métrique de l'espace RVB est très éloignée de la métrique visuelle d'un observateur : 2 points verts très éloignés seront presque confondus pour un observateur, tandis que 2 points bleus à peine distincts dans RVB apparaîtront très différents à un observateur (voir la figure 1.25 à droite).

Pour éviter ces inconvénients, il est possible de passer à des espaces différents dont la représentation se rapproche subjectivement plus de la perception humaine des couleurs, et dans lesquels on peut mieux décorréler les variables.

24. Le métamérisme est l'outil de base de tous les métiers de la couleur chargés d'accorder une teinte à une autre. Exemples : un siège de voiture avec la carrosserie, la restauration d'une peinture ancienne, assortir des boutons à une robe, etc.

25. Lieu des corps noirs = *Planckian locus* en anglais.

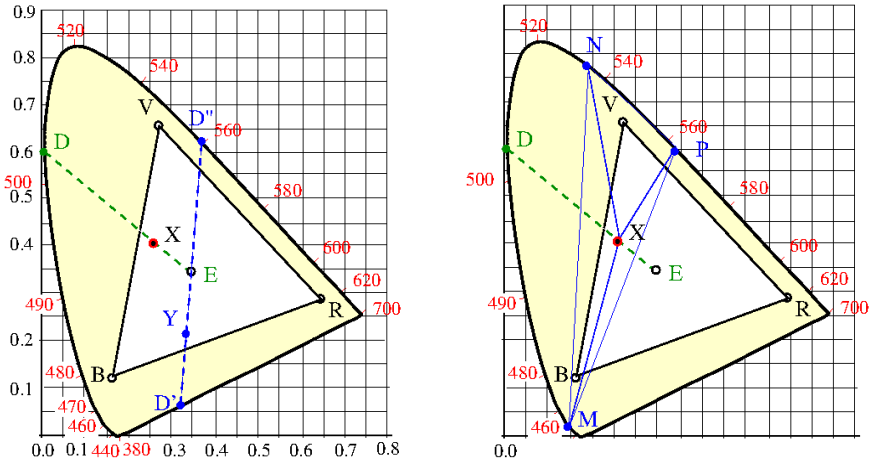

Figure 1.24 – *Espace des couleurs doté de trois primaires RVB et du blanc de référence noté E. A gauche : le point X a pour longueur d'onde dominante un bleu-vert à 505 nm environ (point D). Le point Y, en revanche, n'a pas de longueur d'onde dominante (le point D' correspond à un pourpre qui n'est pas un rayonnement pur). Par convention, on lui attribue la longueur d'onde du complémentaire (point D") à 562 nm. A droite : exemple de métamérisme. Le point X, point de couleur unique en représentation RVB, a pu être créé soit par mélange d'un bleu-vert et de blanc (interpolation linéaire sur la ligne DE), soit par un mélange de bleu (à 460 nm), de vert (à 532 nm) et de rouge (à 564 nm) (interpolation barycentrique dans le triangle MNP), soit encore par une infinité d'autres combinaisons de rayonnements différents. Ils seront tous vus comme une même couleur X par l'observateur.*

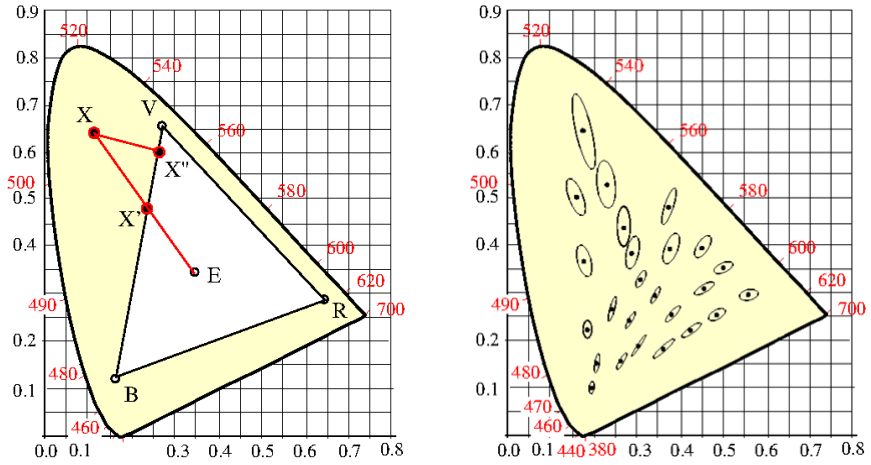

Figure 1.25 – *A gauche : le point X est hors du gamut. Il peut être représenté dans l'espace RVB par le point X" (le plus proche de X) ou par le point X' (même teinte que X, mais désaturée). A droite : Ellipses de MacAdam. Ces ellipses représentent le lieu des points indicernables du point central pour un observateur humain. Ainsi pour les couleurs vertes, un fort déplacement ne s'accompagne d'aucune variation de couleur, tandis que pour les bleus, un très petit changement des valeurs X ou Y est immédiatement perçu par un observateur. Cela interdit à l'utilisateur de prendre une métrique euclidienne dans RVB pour exprimer l'erreur perçue subjectivement par exemple lors d'un codage d'image*

1.5.5 Autres espaces chromatiques

Pour introduire ces différents espaces, il nous faudra également définir, littéralement et mathématiquement, les caractéristiques de la couleur : la teinte, la saturation et le chroma [Fleury and Mathieu, 1962, Seve, 1996], grandeurs bien connues des photographes, mais aussi définir des grandeurs intermédiaires empruntées au sens courant mais qui doivent trouver, en colorimétrie, une référence précise. Nous suivrons les définitions de [Seve, 1996].

1.5.6 L'espace Lab

L'espace Lab, adopté par la CIE en 1976, porte le nom de CIELab. C'est un espace qui essaie d'être uniforme, ce qui signifie que les écarts de couleur dans cet espace (mesurés par la distance euclidienne) sont en première approximation égaux aux écarts de couleur perçus par un observateur.

Il nous faut prendre en compte les 2 composantes du stimulus : sa luminosité et l'information de sa couleur (sa chromaticité).

Cet espace exploite le fait que la luminance d'un rayonnement est indépendante de sa chromaticité.

Pour la luminance on traduit la loi logarithmique de sensibilité de l'œil (loi de Weber-Fechner) par une approximation de par une puissance 1/3, très proche, et par une fonction affine dans les zones sombres.

Pour la chromaticité, on a procédé empiriquement de façon à ramener autant que possible les ellipses de la figure 1.25 à droite à des cercles égaux.

Un point de l'espace Lab est alors décrit par L la clarté, a et b, les coordonnées de chrominance. déduites des composantes X, Y et Z par les équations [26] :

$$\begin{aligned} L &= 116\left(\tfrac{Y}{Yn}\right)^{\frac{1}{3}} - 16 \\ a &= 500\left[\left(\tfrac{X}{Xn}\right)^{\frac{1}{3}} - \left(\tfrac{Y}{Yn}\right)^{\frac{1}{3}}\right] \\ b &= 200\left[\left(\tfrac{Y}{Yn}\right)^{\frac{1}{3}} - \left(\tfrac{Z}{Zn}\right)^{\frac{1}{3}}\right] \end{aligned} \tag{1.9}$$

.

On définit de plus un blanc de référence de coordonnées Xn, Yn, Zn déterminées en fonction de la nature de l'illuminant choisi [Seve, 1996] :

— illuminant A : Xn = 109,85, Yn = 100, Zn = 35,58.
— illuminant C : Xn = 98,07, Yn = 100, Zn = 118,23.
— illuminant D65 : Xn = 95,04, Yn = 100, Zn = 108,88.

Le plan orthogonal à l'axe des clartés est le plan chromatique contenant les deux axes a et b. On remarque que l'axe a porte l'antagonisme vert-rouge et l'axe b l'antagonisme bleu-jaune de façon assez semblable à ce que l'on trouve dans les premières étapes des voies visuelles telles que décrites à la section 1.1.3. [27].

On préfère souvent caractériser la couleur d'un point de couleur par sa teinte T et son chroma C, plutôt que par les valeurs a et b. Pour définir T et C, on passe en coordonnées cylindriques : le chroma est alors la distance du point à l'axe achromatique, la teinte est définie comme l'angle fait par la direction du blanc au point avec l'axe a :

$$C = \sqrt{a^2 + b^2} \tag{1.10}$$

$$T = \arccos \frac{a}{\sqrt{a^2 + b^2}} \qquad \text{si} \quad b > 0 \tag{1.11}$$

$$= 2\pi - \arccos \frac{a}{\sqrt{a^2 + b^2}} \qquad \text{sinon} \tag{1.12}$$

Propriétés de cet espace L'espace Lab est presqu'uniforme, ce qui signifie qu'un écart entre deux couleurs est à peu près égal à l'écart perçu par l'homme. La composante L est décorrélée des

26. Pour des valeurs de l'argument inférieures à 0,008 856, la fonction puissance $x^{\frac{1}{3}}$ est remplacée par le segment de droite : $7,787x + \frac{16}{116}$.

27. Dans l'espace Lab, les variables a et b varient l'une et l'autre de –299 à 300, prenant ainsi 600 valeurs. Un espace noté La*b*, utilise des valeurs de a*, b* variant de –127 à 128, donc codées sur 8 bits. C'est le plus utilisé en traitement des images.

composantes chromatiques, mais malheureusement le chroma dépend de la clarté. Le passage de RVB à Lab entraîne une perte de précision numérique due aux changements successifs d'espaces, en particulier à l'approximation causée par l'exposant $\frac{1}{3}$.

1.5.7 Les limites de la colorimétrie

Il serait très long de lister toutes les limites que rencontrent aujourd'hui les formalismes colorimétriques présentés ci dessus. Nous insisterons cependant sur quelques points majeurs qui relèvent immédiatement de la complexité de la perception humaine telle que nous l'avons présentée dans la première partie de ce chapitre.

Contraste simultané : Les études ci-dessus ont toutes été faites par observation d'une plage isolée sur un fond neutre. Nous avons cependant insisté dans la description de la vision humaine sur l'importance du contexte dans la perception. C'est un phénomène très important dans la perception des couleurs. Un contexte particulier "tire" une couleur vers des teintes qui l'éloignent du contexte. Ces interactions dépendent de très nombreux paramètres et ne sont pas encore bien modélisés malgré de nombreux travaux très intéressants (E. Land [Land, 1993], E. Provenzi [Provenzi, 2017]). La bonne connaissance que nous avons des liaisons entre cellules réceptrices voisines se heurte à la complexité des étages successifs de mise en forme des signaux, en particulier dans le corps genouillé.

Invariance chromatique : La permanence de l'apparence chromatique de zones très familières (visages, vêtements, environnement) est très généralement vérifiée. L'observateur pourra qualifier un visage de "teinte chair" (le qualificatif associé à la couleur des visages des commentateurs de TV en Europe), même dans le cas de systèmes de reproduction très mal réglés. Cette particularité est généralement dénommée par **constance des couleurs**. Elle fait bien sûr appel à notre acquis mémoriel pour compenser une image pauvre ou fortement déformée.

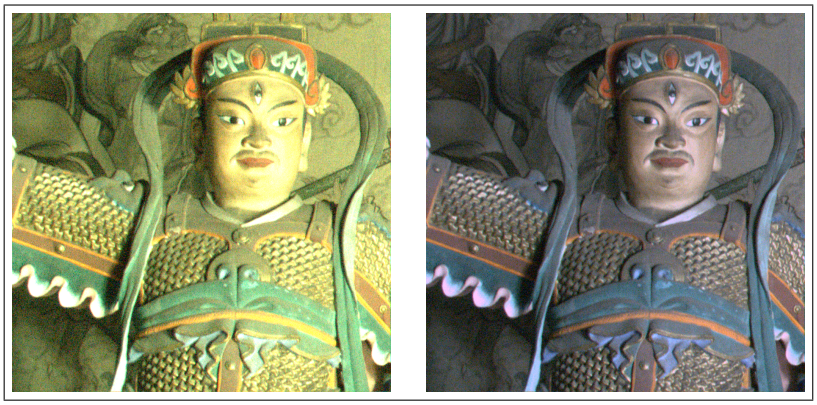

Figure 1.26 – *A gauche, photo prise sous un éclairage fortement coloré. Les teintes représentées ici sont celles que l'on mesurerait objectivement sous cet éclairage. A droite, image corrigée par balance des blancs correspondant assez bien à celle que perçoit un observateur qui observe la scène sous cet éclairage. C'est également une image proche de celle que l'observateur gardera en mémoire.*

Sous des éclairages très colorés, un observateur est très vite sensible aux distorsions chromatiques mais il s'en accomode aisément. Il est alors capable de "voir" blanche une robe objectivement jaune ou verte sous l'effet d'un illuminant violent. Cette capacité du système visuel de faire abstraction (en partie) de l'illuminant pour percevoir la scène comme si elle était vue en éclairage neutre n'est pas du tout expliquée par les modèles de perception précédemment décrits car ils mettent en jeu le contenu mémoriel de notre expérience visuelle. Il est cependant si fort (voir figure 1.26) qu'il amène les développeurs d'applications photographiques à concevoir des systèmes de correction de la

"balance des blancs" de façon que les photos obtenues sous des éclairages très colorés ressemblent plus à la mémoire qu'en gardent les participants à l'expérience qu'à la réalité physique de l'expérience ([Maitre, 2015], chap. 5).

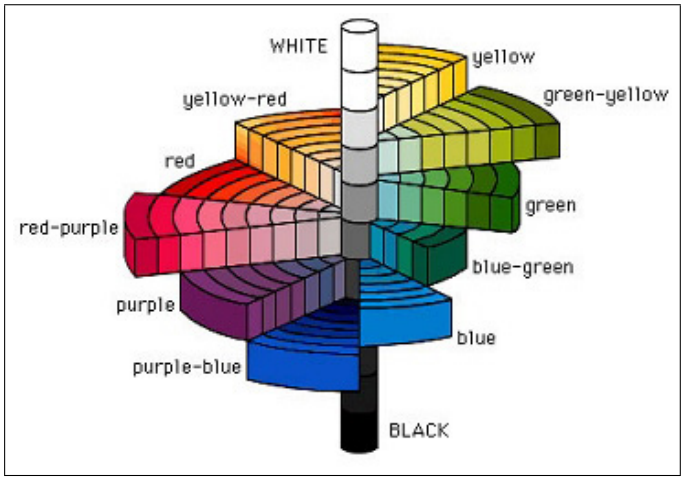

Figure 1.27 – *H. Munsell a développé un nuancier qui présente, dans l'espace tridimensionnel, toutes les couleurs perceptibles à une distance égale les unes des autres. Ce type de nuancier (il en existe de nombreux autres) permet de partager des références colorées en faisant abstraction du repère colorimétrique.*

Harmonie colorée : La juxtaposition de couleurs est souvent affectée d'appréciations qualitatives indiquant que l'association plaît ou déplaît à l'observateur. Ces éléments de goût qui conduisent à des appréciations de la palette d'un peintre ou de l'harmonie d'assemblages chromatiques sont communs, ils ont fait l'objet de travaux académiques (Munsell, Moon et Spencer, Matsuda) qui ont tenté de les raccrocher aux espaces CIE (figure 1.27). Nous ne disposons malheureusement que de très peu éléments pour rattacher ces notions de "plaisir" aux espaces perceptuels [Maitre, 2022].

Bibliographie

[Adelson et al., 1984] Adelson, E., Anderson, C., Bergen, J., Burt, P., and Ogden, J. (1984). Pyramids methods in image processing. *RCA Engineers*, 29(6) :33–41.

[Brown et al., 2007] Brown, G., Perthen, J., Liu, T., and Buxton, R. (2007). A primer on functional magnetic resonance imaging. *Neuorpsychol Rev.*, 17 :107–125.

[Burt, 1984] Burt, P. J. (1984). The pyramid as a structure for efficient computation. In *Multiresolution image processing / analysis*, pages 6–35. Springer Berlin Heidelberg.

[Chalmers, 2010] Chalmers, D. (2010). *L'esprit conscient.* Les Editions d'Ithaque.

[CIE, 2006] CIE (2006). Fundamental chromaticity diagram with physiological axes - part 1. CIE - Vol 170-1.

[Damasio, 1994a] Damasio, A. (1994a). *Descartes' error : emotion, reason and the human brain.* Grosset/Putnam.

[Damasio, 1994b] Damasio, A. R. (1994b). *L'erreur de Descartes.* Odile Jacobs.

[Dennett, 1993] Dennett, D. (1993). *La conscience expliquée.* Odile Jacob, (Paris).

[Desolneux et al., 2008] Desolneux, A., Moisan, L., and Morel, J. (2008). *From Gestalt Theory to Image Analysis : A Probabilistic Approach,*, volume 34. Springer-Verlag, Interdisciplinary Applied Mathematics.

[Faugeras, 1993] Faugeras, O. (1993). *Three-dimensional computer vision : a geometric view-point.* MIT Press.

[Fize, 2004] Fize, D. (2004). *La catégorisation visuelle rapide*, chapter 4 in Imagerie cérébrale fonctionnelle électrique et magnétique (Renault, B. ed), pages 69–93. Lavoisier (Paris).

[Fleury and Mathieu, 1962] Fleury, P. and Mathieu, J. P. (1962). *Images optiques.* Eyrolles, Paris, 3 edition.

[Gaillard, 2024] Gaillard, R. (2024). *L'homme augmenté.* Grasset.

[Goethe, 1792] Goethe, J. (1792). *De l'expérience considérée comme médiatrice entre l'objet et le sujet.* Œuvres complètes, Archives Karéline - 2015.

[Gordon, 1998] Gordon, I. (1998). *Theories of visual perception.* John Wiley.

[Gori, 2018] Gori, P. (2018). *Introduction to Magnetic Resonance Imaging (MRI) - Diffusion and Functional Brain Imaging.*, Telecom-Paris (Cours IMA204) edition.

[Goutcher and Hibbard, 2014] Goutcher, R. and Hibbard, P. (2014). Mechanisms for similarity matching in disparity measurement. *Frontiers in Psychology*, 4 :1014.

[Hubel and Wiesel, 1959] Hubel, D. and Wiesel, T. (1959). Receptive fields of single neurones in the cat's striate cortex. *The Journal of physiology*, 148(3) :574–591.

[Julesz, 1971] Julesz, B. (1971). *Foundations of Cyclopean Perception.* Chicago University Press.

[Koehler, 1929] Koehler, W. (1929). *Psychologie de la forme.* Gallimard, Idées, Paris.

[Kragel et al., 2019] Kragel, P., Reddan, M., LaBar, K., and Wager, T. D. (2019). Emotion schemas are embedded in the human visual system. *Science advances*, 5(7).

[Land, 1993] Land, E. H. (1993). *Edwin H. Land's essays.* McCann, M., Society for Imaging Science and Technology, Springfield, Etats Unis.

[Le Grand, 1964] Le Grand, Y. (1964). *Optique physiologique, tome 1 : la dioptrique de l'oeil et sa correction.* Masson (Paris).

[Lindeberg, 1994] Lindeberg, T. (1994). Scale-space theory. a basic tool for analyzing structures at differentscales. *Journal of Applied Statistics*, 21((1+2)) :225–270.

[Maitre, 2015] Maitre, H. (2015). *Du photon au pixel : l'appareil photographique numérique.* ISTE Editions, Londres, Royaume Uni.

[Maitre, 2022] Maitre, H. (2022). *Esthetique de la photographie numérique.* ISTE Group, London.

[Marr, 1982] Marr, D. (1982). *Vision : A Computational Investigation into the Human Representation and Processing of Visual Information.* W. H. Freeman and Company, 1982.

[Meyer, 1997] Meyer, P. (1997). *L'œil et le cerveau.* Odile Jacob, (Paris).

[Neri, 2005] Neri, P. (2005). A stereoscopic look at visual cortex. *J. Neurophysiology*, 93 :1823–1826.

[Pessoa and Adolphs, 2010] Pessoa, L. and Adolphs, R. (2010). Emotion processing and the amygdala : From a 'low road' to'many roads' of evaluating biological significance. *Nat. Rev. Neurosci.*, 11 :773–783.

[Petitot, 2008] Petitot, J. (2008). *Neurogéométrie de la vision.* Les Editions de l'Ecole Polytechnique.

[Poggio, 1995] Poggio, G. (1995). Mechanisms of stereopsis in monkey visual cortex. *Cerebral Cortex*, 5(3) :193–204.

[Provenzi, 2017] Provenzi, E. (2017). *Computational Color Science : Variational Retinex-like methods.* ISTE-Wiley, London.

[Purves et al., 2015] Purves, D., Augustine, C., Fitzpatrick, D., Hall, W., LaMantia, A., and White, L. (2015). *Neurosciences.* De Boeck.

[Schopenhauer, 1966] Schopenhauer, A. (1966). *Le monde comme volonté et comme représentation.* PUF (Paris).

[Schupp et al., 2004] Schupp, H., Cuthbert, B., Bradley, M., Hillman, C., Hamm, A., and Lang, P. (2004). Brain processes in emotional perception : Motivated attention. *Cognition and emotion*, 18(5) :593–611.

[Seve, 1996] Seve, R. (1996). *Physique de la couleur : de l'apparence colorée à la technique colorimétrique.* Masson - Physique fondamentale et appliquée.

[Thorpe et al., 1996] Thorpe, S., Fize, D., and Marlot, C. (1996). Speed of processing in the human visual system. *Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. nature,*, 381(6582) :520–521.

[Vienot and Le Rohellec, 2012] Vienot, F. and Le Rohellec, J. (2012). Colorimétrie et physiologie : la spécification LMS. In Fernandez-Maloigne, C., Robert-Inacio, F., and Macaire, L., editors, *Couleur numérique : acquisition, perception, codage et rendu*, chapter 1, pages 19–40. Hermès-Lavoisier, Paris.

[Wandell, 1995] Wandell, B. (1995). *Foundations of Vision : Behavior, Neuroscience and Computation.* Sinauer Ass.

[Yarbus, 1967] Yarbus, A. (1967). *Eye movements and vision.* Plenum Press.