

Contents

1	Interaction paradigms	1
1.1	Command lines	2
1.2	WIMP (Windows, icons, menus and pointers)	2
1.3	Direct manipulation interface	3
1.4	Form filling	3
1.5	Experiment with the four previous paradigms	3
1.6	Virtual reality	4
1.7	Tangible interaction	5
1.8	Augmented reality	5
1.9	Difference between virtual reality and augmented reality	5
1.10	Voice command devices	5
1.11	Gesture control	6
1.12	Braincomputer interface	6
1.13	Bionic interfaces	6
1.14	Wearables interfaces	6
1.15	Interaction ambient	7
1.16	Final thoughts	7
1.17	Exercises	8
2	Introduction to Computer Graphics	8
2.1	Modeling	9
2.2	Rendering	10
2.2.1	Visibility	10
2.2.2	Shading and materials	12
2.3	Exercises	13
3	Vision humaine	14
3.1	Structure de l'oeil	14
3.2	Effets visuels	17
3.2.1	L'effet Stroop	17
3.2.2	Perception du mouvement	18
3.2.3	Perception de l'amorçage	19
3.2.4	Perception de la profondeur	21
3.2.5	Recrer la profondeur en 2D	22
3.3	Remarques	22
3.4	Exercises	23

1 Interaction paradigms

The term Human-Computer Interaction (HCI) refers to the study of interaction between humans and computers. The scientific field of HCI has seen considerable advancements in the last 30 years, especially since the

publication of Card, Moran and Newells Psychology of Human Computer Interaction in 1983. Most of the research efforts in HCI attempt to make use of human skills and abilities that are naturally developed by leading a life in this physical world. One can classify different paradigms of interaction that take place between humans and computers:

1.1 Command lines

Initial interfaces were command-line interfaces where a user sits in front of a terminal screen, and enters a line specific command to perform specific tasks and wait for a reply. Only a part of the terminal screen real estate was used. Some of the concepts that were introduced were the concept of infinite loop (prompt), the immediate execution of commands (enter), a syntax allowing obligatory or optional parameters, the maintenance of a history of commands and the introduction of keyboard short-cuts. They were further improved with the introduction of textual interfaces that made use of the entire terminal screen.

1.2 WIMP (Windows, icons, menus and pointers)

The WIMP paradigm relies on Windows, Icons, Menus and Pointers to interact with the user.

- A window runs a self-contained program, isolated from other programs that (if in a multi-programmed operating system) run at the same time in other windows.
- An icon acts as a shortcut to an action the computer performs (e.g., execute a program or task).
- A menu is a text or icon-based selection system that selects and executes programs or tasks.
- The pointer is an on-screen symbol that represents movement of a physical device that the user controls to select icons, data elements, etc.

Post-WIMP comprises work on user interfaces, mostly graphical user interfaces, which attempt to go beyond this paradigm.

The reason WIMP interfaces have become so prevalent is that they are very good at abstracting work-spaces, documents and their actions. Their analogous paradigm to documents as paper sheets or folders makes WIMP interfaces easy to introduce to other users.

However, WIMP interfaces are not optimal for working with complex tasks such as showing 3D models (computer-aided design), working on large amounts of data simultaneously (interactive games) or simply portraying

an interaction for which there is no defined standard widget. WIMPs are usually pixel-hungry, so given limited screen real estate they can distract attention from the task at hand.

1.3 Direct manipulation interface

It was introduced in the context of the creation of office applications and the desktop metaphor. It involves continuous representation of objects providing actions that are rapid, reversible, incremental and with feedback. These actions should correspond to manipulation of physical objects. Example: resizing a graphical shape by dragging its corners or edges with a mouse or doing a drag-n-drop to delete a file.

Benefits: can make it easier for a user to learn and use an interface, having rapid, incremental feedback allows a user to make fewer errors and complete tasks in less time.

1.4 Form filling

It is preferred for bureaucratic and routine tasks. Introduces the notion of obligatory, optional or conditional fields, the notion of unfolding menus, the guide of the user during the process and the use of short-cuts. It allows the automatic verification of inputs. Example: online reservation of a flight.

1.5 Experiment with the four previous paradigms

The experiment that can be found <http://cs211labs.epfl.ch> does a simulation of the purchase of five tickets in each of the four previous interaction modes. The conclusion is that the best paradigm depends on the task, the user and the details of the design.

Depending of the degree of use of the tool we can have a beginner, an intermediate or an expert user. A beginner user will only retain the aspects of the task that he needs to know in the real world without using a computer. The intermediate user will also know more advanced features that help to do the translation of the task from the real world to the computer. Finally, the expert user master aspects inherent to the computer program itself that don't have to do with the task itself.

Example: if we take OpenOffice Word, the beginner should know that he needs an index, an introduction and a body for writing an essay. The intermediate user will know that in order to structure the body of his essay he needs to structure the text in paragraphs and can set the parameters to separate paragraphs and set its style. Finally, the expert user will master where are the options set in the application or what is the formal languages used for specific actions.

1.6 Virtual reality

Computer interactive simulation using sound and visual effects corresponding to real, imaginary or semi-imaginary environments.

Examples: Minecraft (characters represented by avatars although the world is not real), eye surgery learning.

An interactive computer simulation is immersive if the user receives similar stimulus to the ones he would receive in the environment that is being simulated. The stimulus can be in all senses: sight, smell, touch or sound.

Examples: a cave automatic virtual environment (CAVE) is an immersive virtual reality environment where video projectors are directed to between three and six of the walls of a room-sized cube. CAVE is typically a video theater situated within a larger room. The user wears 3D glasses inside the CAVE to see 3D graphics generated by it. People using the CAVE can see objects apparently floating in the air, and can walk around them, getting a proper view of what they would look like in reality. A CAVE user's movements are tracked by the sensors typically attached to the 3D glasses and the video continually adjusts to retain the viewers perspective.

A question remains to be answered. Do these virtual reality experiences really influence the users as the real world does. Several experiments have been conducted in this sense, measuring physiological and cognitive responses. For instance, virtual reality users hearts have pumped intensely while crossing a virtual pit (walking through a piece of wood surrounded by virtual cliffs).

Difference haptics (related to the touch sense) devices has been implemented. They give the sensation of touch by applying forces (force feedback) or vibrations to the movements of users. This needs the presence of sensors for measuring the forces applied by the user. The user can then feel the specific resistance, elasticity or rugosity of the surface

Applications: surgery simulations, bass shakers vibrations in cinemas.

A particularly interesting application is a data glove or wired glove. It is an input device for human-computer interaction worn like a glove. Various sensor technologies are used to capture physical data such as bending of fingers. Often a motion tracker, such as a magnetic tracking device or inertial tracking device, is attached to capture the global position/rotation data of the glove.

These movements are then interpreted by the software that accompanies the glove, so any one movement can mean any number of things. Gestures can then be categorized into useful information, such as to recognize sign language or other symbolic functions.

Expensive high-end wired gloves can also provide haptic feedback, which is a simulation of the sense of touch. This allows a wired glove to also be used as an output device. Wired gloves are often used in virtual reality environments and to mimic human hand movement by robots.

1.7 Tangible interaction

Tangible Interaction has come to be the 'umbrella term' used to describe a set of related research and design approaches which have emerged in several disciplines. It covers user interfaces and interaction approaches that emphasize:

- tangibility and materiality of the interface which is then connected with sensors with the computer. These sensors can be in the form of cameras, touch-tables or RFID.
- the objects employed are specific to the task at hand which can be the visit to a museum to learn about a specific topic or directed to children to learn programming. It is very usual that these physical environments are thought to be multi-user to enhance teamwork.

1.8 Augmented reality

Augmented reality (AR), is a live direct or indirect view of a physical, real-world environment whose elements are augmented by computer-generated sensory input such as sound, video, graphics or GPS data. Augmented reality enhances one's current perception of reality, whereas in contrast, virtual reality replaces the real world with a simulated one. Augmentation techniques are typically performed in real-time, that is, there are time limits that must be met, and in a semantic context with the addition to the environmental elements of overlaying supplemental information like scores over a live video feed of a sporting event.

There are several hardware instruments that can be used to this goal. Using transparencies such as glasses or projections. We can even have interactive surfaces where we should ask ourselves whether it is better to project from the upside or from the downside. A special characteristic of these surfaces is whether they are multi-touch or not. Multi-touch means that we can recognize the presence of more than one or more than two points of contact with the surface.

1.9 Difference between virtual reality and augmented reality

Virtual reality offers a digital recreation of a real life setting, while augmented reality delivers virtual elements as an overlay to the real world.

1.10 Voice command devices

These are interfaces that allow the user to communicate using its voice with the computer. A popular system for this is Siri. However, there is still a lot to do in human voice recognition to generalize these tools. The level of precision depends on factors like the size of the vocabulary of the user, the

environmental noise, the time that a particular user has been using it since Siri claims that it can learn with time or certain others semantic problems. There are also certain challenges in its usability.

1.11 Gesture control

This interface is based in the recognition of human gestures. In EPFL, researcher Frdric Kaplan has developed such a system to simulate tennis games. Microsoft Kinect is also an examples. It enables users to control and interact with their console/computer without the need for a game controller, through a natural user interface using gestures and spoken commands.

1.12 Braincomputer interface

These interfaces allow the communication of the user's brain activity to the computer. In EPFL, researchers have obtained results controlling a wheelchair. The thoughts of the user activate specific brain patterns that are recorded by electroencephalography (EEG) using a helmet with electrodes. These patterns are then interpreted by a computer that transmits a command to the chair. Interestingly, it takes some time to the user of this systems to adapt their thoughts to the machines they use. It is a process of mutual apprenticeship between human and machine.

1.13 Bionic interfaces

Bionics is the application of biological methods and systems found in nature to the study and design of engineering systems and modern technology. Therefore a bionic interface is an interface that profits of this knowledge to communicate the user and the computer.

An interesting experiment was carried out in EPFL by the team of Silvestro Mcera to make a bionic hand with which patients could adjust their force to take objects and identify their shape and texture. The prosthesis contained some sensors capable of reacting to the tension of artificial tendons at translating them into electric impulses. This electric signals could then be transmitted to the actual external nerves of the patient arm.

1.14 Wearables interfaces

Wearable technology or wearables are electronic devices with microcontrollers that can be worn on the body as implants or accessories. The designs often incorporate practical functions and features.

This can be used for instance for sousveillance, which is the recording of an activity by a participant in it by means of wearing certain objects. It has also been used in the fashion industry. CuteCircuit was the first fashion

company offering smart textile-based garments that create an emotional experience for their wearers using smart textiles and micro electronics.

1.15 Interaction ambient

The notion of ambient devices revolves around core concept of immediate access to information. The original developers of the idea (HYATT, ROSE, 2002) state that in the majority of cases an ambient device is designed to provide support to people in carrying out their everyday activities in an easy and natural way. An average person living in a modern society is being overloaded with abundance of information on a daily basis. Through the introduction of ambient devices into their day-to-day routine an individual gains an opportunity to decrease the amount of effort to process incoming data, thus rendering self more informed and productive (ROSE, 2002).

The key issue lies within taking Internet-based content (e.g. traffic congestion, weather condition, stock market quotes) and mapping it into a single, usually one-dimensional spectrum (e.g. angle, colour). According to one of the concept originators David L. Rose this way the data is represented to an end user seamlessly, and its procurement requires an insignificant amount of cognitive load.

One of the examples of the ambient device technology is Ambient Orb, introduced by Ambient Devices in 2002 (KIRSNER, 2002). The device itself was a glowing sphere which was continuously displaying data through perpetual changes in colour. Ambient Orb was customizable in terms of content and its subsequent visual representation. For instance, when the device was set to monitor a particular stock market index (e.g. NASDAQ), the Orb glowed green/red to represent the upward/downward movement of the stock prices; alternatively, it turned amber when the index is unchanged. Nabeel Hyatt stated that the device was marketed as an interior design item with additional functionality.

1.16 Final thoughts

Often, interfaces are said to be intuitive. However, we all have to learn how things work. The fact that we think that an interface is intuitive is more related with the fact that we forget that we have learnt it. To properly assess the quality of a good interface one has to measure:

- the learning curve
- the time needed to complete tasks
- the number of errors
- the user satisfaction

1.17 Exercises

Exercise 1.1.

Suppose a website for reserving plane tickets. Classify the following knowledges into the different categories:

- If a user waits very long before confirming his choice the session is stopped and he has to restart all the process. (related to the computer translation)
- A user has to enter the departure and arrival airports that are in the scope of the company's flights. (related to the computer translation)
- Each airport of the world has a code name in three letters. For instance, GVA for Geneva. (related to the task)
- In Swiss.com if the user wants to reserve a ticket in business class, he has to click on "more options" in the first dialogue window. (related to syntax)
- I can order the possible flights in terms of the number of layovers or the price. (related to the computer translation)

Exercise 1.2.

A software program for 3D rendering offers multiple functionalities to architects. What interaction paradigm is the most appropriate for the listed tasks. (It is possible to have several).

- Position a numeric image of the furniture and the equipment in the interior of the future kitchen. (direct manipulation)
- position a numeric image of the future house on a real image of the ground to understand its integration in the landscape. (augmented reality)
- Make the the listing of the required windows for each room of the house. (form filling)
- Propose to the clients an immersive visit in their future house. (virtual reality)

2 Introduction to Computer Graphics

Although computer graphics is a vast field that encompasses almost any graphical aspect, we are mainly interested in the generation of images of 3-dimensional scenes. Computer imagery has applications for film special effects, simulation and training, games, medical imagery, flying logos, etc.

Computer graphics relies on an internal **model** (modeling) of the scene, that is, a mathematical representation suitable for graphical computations. The model describes the 3D:

- shapes
- layout
- materials

This 3D representation then has to be projected to compute a 2D image from a given viewpoint, this is the **rendering** step. Rendering involves:

- projecting the objects (perspective)
- handling visibility (which parts of objects are hidden)
- computing their appearance and lighting interactions

Finally, for animated sequence, the motion of objects has to be specified. We will not discuss animation in this document.

2.1 Modeling

We introduce how the geometry of the scene is represented in the memory of the computer.

The most classical method for modeling 3D geometry is the use of polygons. An object is approximated by a polygonal mesh, that is a set of connected polygons. Most of the time, triangles are used for simplicity and generality.

Each polygon or triangle can be described by the 3D coordinates of its list of vertices. The obvious limitation of triangles is that they produce a flat and geometric appearance. However, techniques called smoothing or interpolation can greatly improve this.

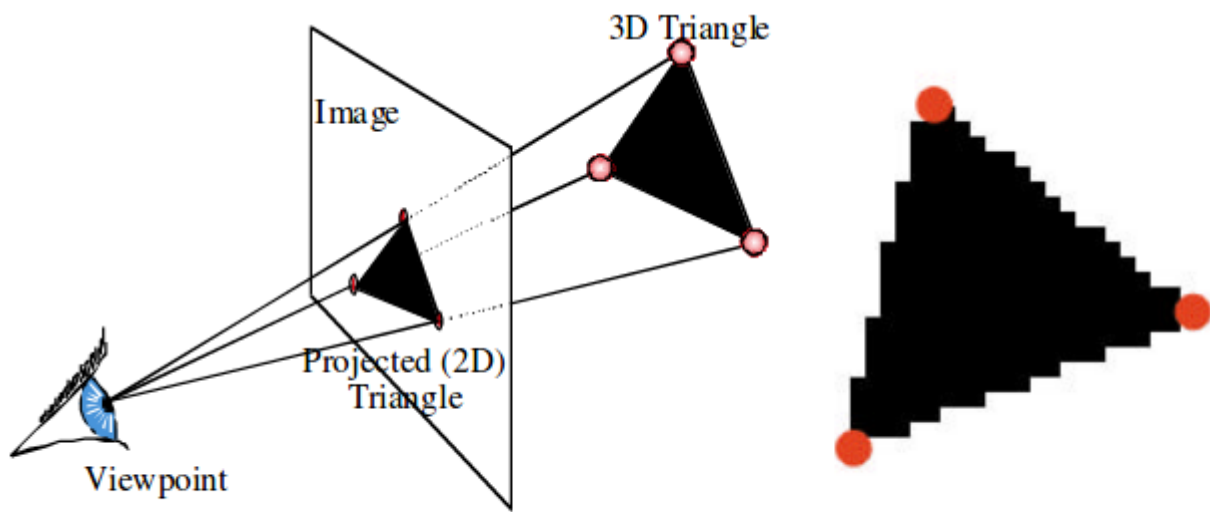
The most classical geometric entities can be directly used as primitives, e.g. cubes, cylinders, spheres and cones. A sphere for example can be simply described by the coordinates of its center and its radius.

More complex mathematical entities permit the representation of complex smooth objects. Spline patches and NURBS are the most popular. They are however harder to manipulate since one does not directly control the surface but so called control points that are only indirectly related to the final shape. Moreover, obtaining smooth junctions between different patches can be problematic. However, the recently popular subdivision surfaces overcome this limitation. They offer the best of both worlds and provide the simplicity of polygons and the smoothness of patches.

2.2 Rendering

The image projection of the 3D objects is computed using linear perspective. Given the position of the viewpoint and some camera parameters (e.g. field of view), it is very easy to compute the projection of a 3D point onto the 2D image. For mathematics enthusiasts, this can be simply expressed using a 4×4 matrix.

In most methods, the geometric entities are then **rasterized**. It consists in drawing all the pixels covered by the entity. In the example below, the projections of the 3 red points have been computed using linear perspective, and the triangle has then been rasterized by filling the pixels in black.



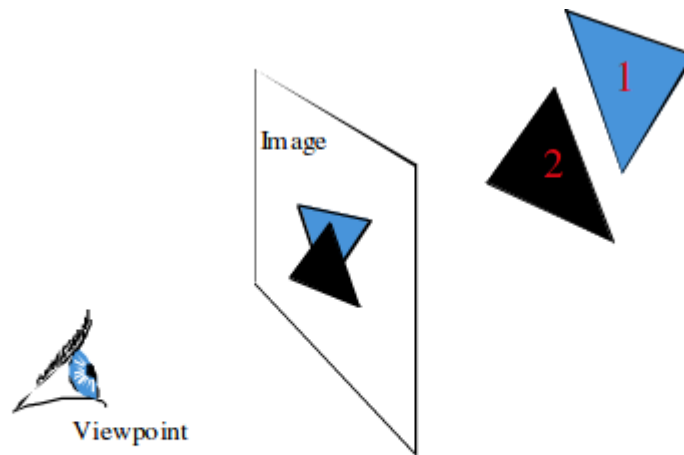
Projection (left) and rasterization (right) of a triangle.

For richer rendering, the color of each rasterized pixel must take into account the optical properties of the object, as we will discuss below.

2.2.1 Visibility

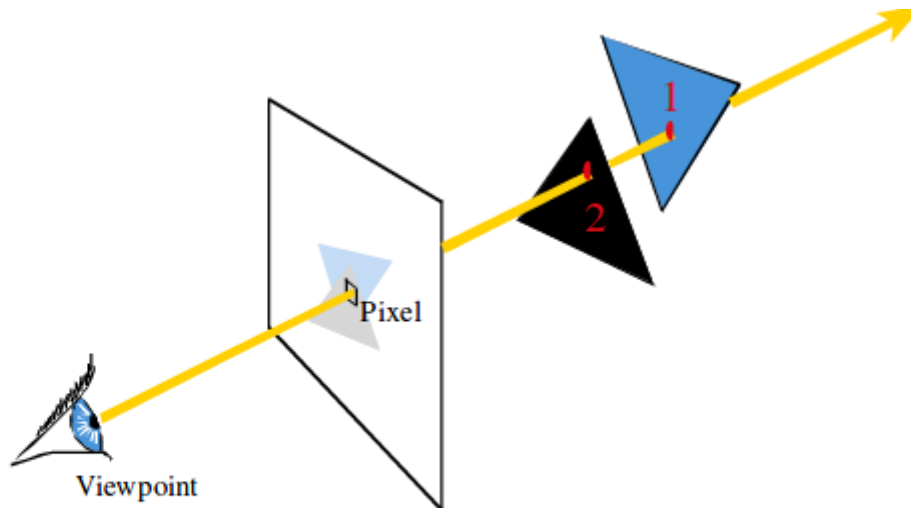
If the scene contains more than one object, occlusions may occur. That is, some objects may be hidden by others. Only visible objects should be represented. Visibility techniques deal with this issue. One classical algorithm that solves the visibility problem is the so-called painters algorithm.

It consists in sorting the objects or polygons from back to front and rasterizing them in this order. This way, front-most polygons cover the more distant polygons that they hide.



The Painter's algorithm. Triangle 1 is drawn first because it is more distant. Triangle 2 is drawn next and covers Triangle 1, which yields correct occlusion.

The **ray-tracing** algorithm does not use a rasterization phase. It sends one ray from the eye and through each pixel of the image. The intersection between this ray and the objects of the scene is computed, and only the closest intersection is considered.



Ray-tracing. A ray is sent from the eye and through the pixel. Since the intersection with 2 is closer than the intersection with 1, the pixel is black.

The z-buffer method is the most common nowadays (e.g. for computer graphics cards). It stores the depth (z) of each pixel. When a new polygon

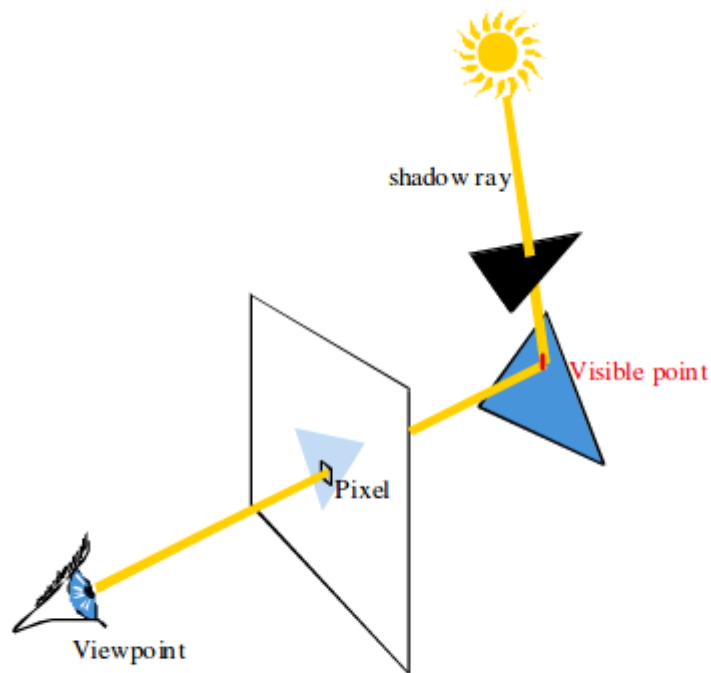
is rasterized, for each pixel, the algorithm compares the depth of the current polygon and the depth of the pixel. If the new polygon has a closer depth, the color and depth of the pixel are updated. Otherwise, it means that for this pixel, a formerly drawn polygon hides the current polygon.

2.2.2 Shading and materials

Augmenting the scene with light sources allows for better rendering. The objects can be shaded according to their interaction with light. Various shading models have been proposed in the literature. They describe how light is reflected by object, depending on the relative orientation of the surface, light source and viewpoint.

Texture mapping uses 2D images that are mapped on the 3D models to improve their appearance.

Shading and material models only take into account the local interaction of surfaces and light. They do not simulate shadows that are harder to handle because they imply long-range interactions. A shadow is caused by the occlusion of light by one object. Ray-tracing, for example, can handle shadows, but requires a shadow computation for each pixel and each light source. A shadow ray is sent from the visible point to the light source. If the ray intersects an object, then the visible point is in shadow.



Shadow computation using ray-tracing.
The visible point is in shadow because the black triangle occludes the light source.

More complex lighting interactions can then be simulated. In particular, objects that are illuminated by a primary light source reflect light and produce indirect lighting. This is particularly important for indoor scenes. Global lighting methods take into account all light inter-reflections within the scene.

Rendering can be classified in three categories:

- Non-photorealistic: rendering of scenes in an artistic style, intended to look like a painting or drawing.
- Photorealistic: rendering of scenes intended to look like indistinguishable from a photo.
- Physically based: rendering in a way that more accurately models the flow of light in the real world.

See: http://people.csail.mit.edu/fredo/Depiction/1_Introduction/reviewGraphics.pdf

2.3 Exercises

Exercise 2.1.

What of the following matrices corresponds to the 2D transformation: translation of an homogeneous vector $(-1, 1, 1)$ followed by a rotation of angle $\theta = 270^\circ$.

- We have to multiply matrices $((0, -1, 0), (1, 0, 0), (0, 0, 1))$ and $((0, -1, 1), (1, 0, 2), (0, 0, 1))$.
- The answer is $((0, -1, 1), (1, 0, 2), (0, 0, 1))$.

Exercise 2.2.

What of the following equivalences between points in homogeneous coordinates are true?

- $(-1, -1, -1) \sim (1, 1, 1)$ true, take $\lambda = -1$.
- $(-1, 1, 1) \sim (1, -1, 1)$ false, no λ possible.
- $(-2, 2, -1) \sim (1, -1, 1)$ false, no λ possible.
- $(-1, 1, 0) \sim (1, -1, 0)$ true, take $\lambda = -1$.

3 Vision humaine

3.1 Structure de l'oeil

Ce cours étudie donc la boucle de traitement des images qui part de l'écran, est analysée par l'oeil et le cerveau et génère des actions qui sont perçues par l'ordinateur. Nous nous concentrons ici sur la perception visuelle.

Le point essentiel de ce cours est de montrer qu'une grande partie de la perception visuelle se passe dans le cerveau. C'est le cas de la construction des couleurs, comme nous le verrons plus loin, mais aussi de certaines illusions visuelles, comme la perception du triangle de Nakizsa: les bords du triangle ne sont pas dans l'image mais construits par le cerveau.

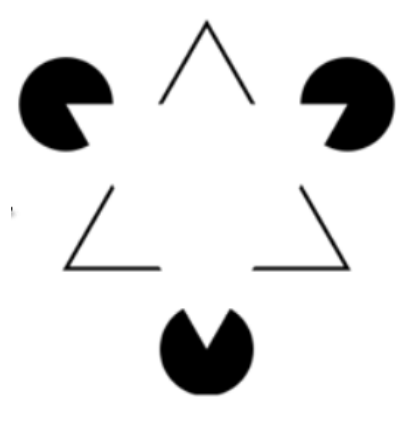
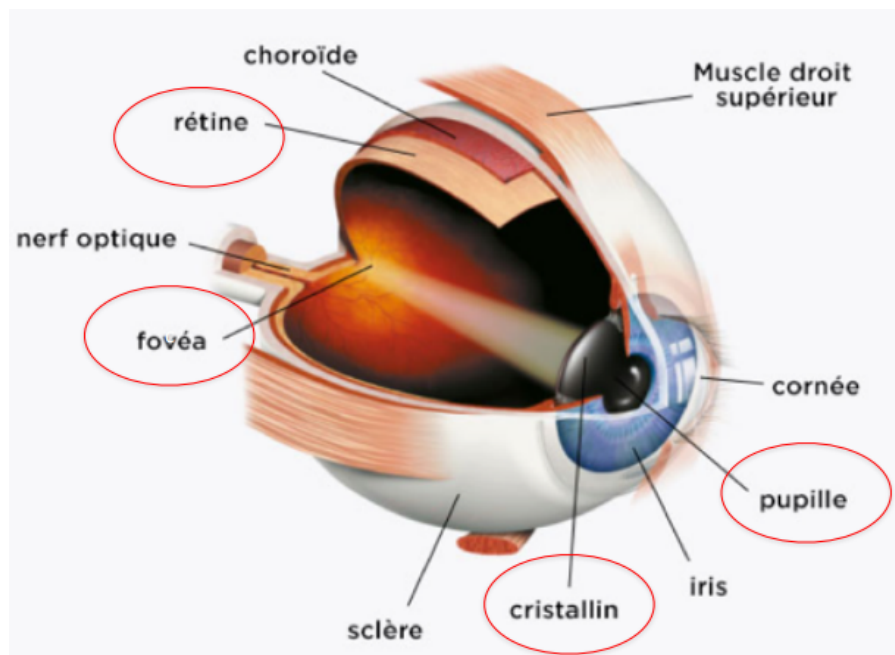


Figure 1: Triangle de Nakizsa

Voici la structure d'un oeil dont 4 éléments sont particulièrement pertinents pour le cours Visual Computing.



- pupille: contrôle la quantité de lumière qui entre dans l'oeil. Sa taille est contrôlée par des mouvements des muscles de l'iris qui agissent par réflexe à la quantité de lumière.
- cristallin: c'est une lentille bi-convexe qui concentre plus ou moins les rayons lumineux sur la rétine. il permet de garder une image nette à des profondeurs de champs différents. Les muscles autour du cristallin l'étirent lorsque l'objet regardé s'éloigne, un phénomène appelé accommodation.
- rétine: sorte d'écran placé au fond de l'oeil. Elle est la partie de l'oeil qui capte les photons grâce à un grand nombre de photorécepteurs. Elle couvre le 75 % de la surface interne de l'oeil.
- macula et fovéa: la macula est la partie centrale de la rétine située dans l'axe de la rétine. La fovéa est une partie de la macula, une petite dépression qui correspond à la zone qui nous donne la plus grande activité visuelle. Elle comprend une très forte densité de certains photorécepteurs appelés cônes alors que le reste de la macula comporte davantage des récepteurs appelés bâtonnets.

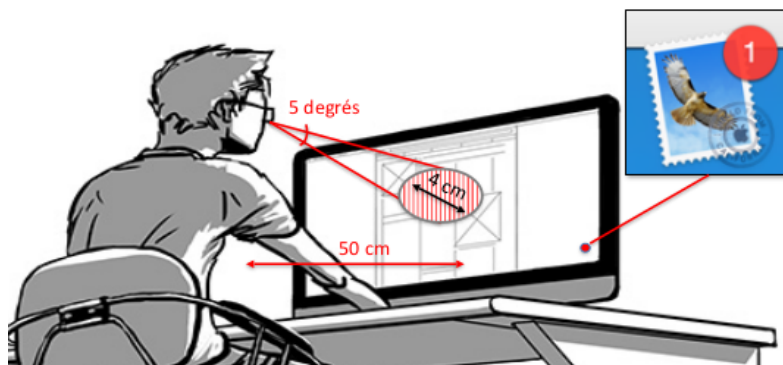
Les bâtonnets et les cônes sont des cellules photosensibles. Les bâtonnets ne perçoivent pas les couleurs mais sont plus sensibles aux niveaux de gris. Ils permettent notamment la vision par faible luminosité. Ils semblent aussi impliqués dans la vision périphérique pour la perception du mouvement. Environ la rétine contient 100 millions.

Les cônes se situent dans la fovéa. Ils sont environ 6 millions. Ils ont besoin d'avantage de lumière et sont responsables de la perception de la couleur. En réalité, les cônes ne reconnaissent pas la couleur mais nous possédons 3 types de cônes sensibles à des longueurs d'onde différentes. Les cônes n'envoient pas un message de couleur au cerveau mais lui communiquent le nombre des photons qui les ont heurtés. Le daltonisme résulte du fait que certains types de cônes ne possèdent pas les pigments photosensibles qui captent les photons.

Il existe cependant, dans chaque oeil, une zone dépourvue de récepteurs car c'est le point d'ancrage du nerf optique qui conduit les informations au cerveau. Il s'appelle le **point aveugle**: tout objet situé dans l'axe rétine-point aveugle sera invisible pour cet oeil. Si nous avons les deux yeux ouverts, l'objet sera toujours visible par un oeil.

En classe, nous font l'expérience suivante. Fermez l'oeil droit. Regardez le losange rouge avec l'oeil gauche. Le professeur traverse l'écran doucement. Vous percevez le professeur en vision périphérique mais il va disparaître un moment.

La structure de la rétine définit notre acuité visuelle dans les différentes zones de notre champ de vision. Notre vision la plus précise se limite à un angle de 5 degrés. Néanmoins notre oeil est sensible aux mouvements périphériques. Il ne glisse pas de manière continue sur le texte mais l'explore par saccades (voir eye tracking). L'oeil se pose pendant des courtes pauses ou fixations d'un quart de seconde. Puis, il se déplace très rapidement vers un autre endroit, ce qu'on appelle une saccade. L'oeil prélève de l'information quand il fait une fixation alors qu'il est aveugle pendant la saccade.



Quelle quantité d'information visuelle est-elle comprise dans cette zone fovéale de 4 cm de diamètre. Cela dépend de 6 paramètres énumérés dans les transparents suivants:

- l'acuité visuelle de l'utilisateur, de sa distance, de ses lunettes !
- zoom choisi par l'utilisateur !

- dimensions de l'écran, définition de l'écran, résolution, profondeur.

Le nombre de pixels par écran n'a pas cessé d'augmenter....

Les 3 types de cônes envoient chacun un nombre de stimulations électriques en fonction du nombre de photons qui les ont heurté. C'est donc le cerveau humain, notamment le cortex visuel, qui intègre ces 3 informations pour déterminer la couleur.

Il semblerait que notre cerveau soit capable de distinguer 10 millions de couleurs. attention, il ne s'agit pas de leur donner un nom mais de détecter que deux couleurs sont différentes. Le nombre de 10 millions dépend donc du protocole expérimental utilisé, ce qui explique qu'il n'y ait pas consensus. Mais cela signifierait qu'un écran qui a une profondeur de pixel de 24 bits peut générer d'avantage des couleurs que ce que l'oeil humain peut discriminer.

Dans quelques images on peut voir que la même différence en pourcentage de cyande est ou non est perceptible. Il s'explique par la faible précision du système.

La taille d'un fichier image dépend donc du nombre des points dans cette image ainsi que du nombre des bits d'information par point, c'est à dire la profondeur du pixel. Heureusement, ce nombre est ensuite diminué par des méthodes de compression d'image. D'autres cours de l'EPFL traitent abondamment des méthodes de compression. Après une compression "lossless", la décompression du fichier permet de retrouver exactement chaque bit d'information du fichier original. C'est le cas de fichiers .ZIP. Dans une méthode "lossy", ce n'est pas le cas. Par exemple, si on prend la moyenne de couleur entre pixels voisins, un pixel blanc et un noir deviendront deux pixels gris après la décompression on pourra déterminer s'ils s'agissait au départ de deux gris ou d'un blanc et d'un noir. La question est de savoir combien d'information peut être perdu à la décompression sans que l'oeil humain ne s'en aperçoive. En ce qui concerne les images, les standards PNG et GIF sont "lossless" alors que JPEG est "lossy".

3.2 Effets visuels

3.2.1 L'effet Stroop

Les transparents qui précèdent concernaient la capacité de distinguer deux couleurs. Il ne s'agissait pas de les nommer, ce qui est une tâche cognitive (trouver le nom dans sa mémoire à long-terme). Voici une expérience qui montre les interférences entre la perception et la cognition.

Rouge

Vert

Monsieur Stroop a découvert un effet intéressant lié à la tâche suivante: le sujet doit lire un nom de couleur écrit soit dans la couleur du nom soit dans une autre couleur. Il faut nommer la couleur du mot. L'expérience au CHILI est un peu différente: il faut dire si l'énoncé est correct.

Le temps nécessaire pour énoncer la couleur d'un mot est plus élevée si le mot désigne une autre couleur. Ce phénomène d'interférence résulte du fait qu'on ne puisse s'empêcher de lire le mot, de lui donner une valeur sémantique. La perception ne se fait de manière indépendante des autres activités du cerveau, comme nous allons continuer à le démontrer.

Comme illustré au début de ce cours, la couleur n'est pas perçue par l'oeil mais reconstruite par le cortex visuel en fonction du nombre de signaux reçus par les trois types de cônes. C'est aussi dans le cerveau que se construisent la perception du mouvement, les effets d'amorçage et la perception de la profondeur.

3.2.2 Perception du mouvement

La perception du mouvement doit forcément résulter de la succession d'images, celles-ci ne sont pas gardées dans l'oeil mais envoyées instantanément au cerveau. Les avis divergent sur le nombre d'images qu'un oeil peut percevoir par secondes. Il semblerait que la rétine puisse capter plus de 100 images par secondes (Entre 75 et 150), certains parlent de 1000 fps, le point est controversé. Alors pourquoi le cinéma s'est-il fixé sur le standard de 24 images par secondes? Parce qu'avec 24 images par secondes, voire à partir de 16, nous percevons les mouvements comme étant continus. Ce nombre de 24 ne repose pas sur une réalité physiologique mais sur un standard établi aux premières heures du cinéma. Aujourd'hui certaines TV haute résolution proposent 120 images par secondes voire 240 pour les écrans LEDs, on parle même de TV à 480 fps. Le débat consiste à savoir si ces fréquences offrent une différence perceptible de confort visuel.

Pourquoi perçoit-on des images qui se suivent comme un mouvement continu?

Hypothèse 1: persistance rétinienne.

La première hypothèse, aujourd'hui critiquée par les scientifiques, mais toujours répandue dans l'opinion publique, serait que la rétine garde un moment l'information avant que les cellules photoréceptrices qui ont été

excitées par un photon ne retournent à leur état de repos.

L'image resterait environ entre $1/12$ et $1/25$ de seconde "imprimée" dans la rétine, donc, dès 24 images par seconde, on percevrait une continuité. Le film du cheval ne comporte que 12 images par seconde, on voit que le mouvement n'est pas fluide mais on reçoit néanmoins le mouvement. L'hypothèse de la persistance rétinienne est aujourd'hui abandonnée au profit de l'hypothèse suivante.

Hypothèse 2: effet phi et effet beta.

L'effet phi est la sensation visuelle de mouvement provoquée par l'apparition d'images perçues successives, susceptibles d'être raccordées par un déplacement ou une transformation. C'est le cerveau qui "comblerait" les images intermédiaires, qui imagine le mouvement. L'effet beta est la même illusion de mouvement créé par des images statiques, alors que l'effet phi repose sur des images qui clignent.

En résumé, les hypothèses qui attribuent la construction de la continuité du mouvement au cerveau dominant aujourd'hui celles qui l'attribuent à l'oeil. Ce phénomène qui expliquerait le mouvement apparent dans les "flip-books" et "thaumatropes".

3.2.3 Perception de l'amorçage

La perception est typiquement en mécanisme bottom-up: je perçois l'image d'un objet et je l'associe aux objets que je connais déjà. On parle de processus-bottom up: l'image de l'objet s'imprime sur la rétine qui transmet au cortex visuel lequel communique avec les autres composantes du système cognitif.

Ce texte illustre le processus **bottom-up**:

Le fait qu'un élément visuel brise la régularité du texte. Rome étant écrit en rouge, permet de le sélectionner plus rapidement. Le fait que Rotterdam se mette à trembler attire notre attention. Le fait que Moscou se dépala attire notre attention.

Parmi l'ensemble des stimuli présentés, certains ont des propriétés qui attirent notre attention, celui-ci brise une régularité établie par l'ensemble du stimulus.

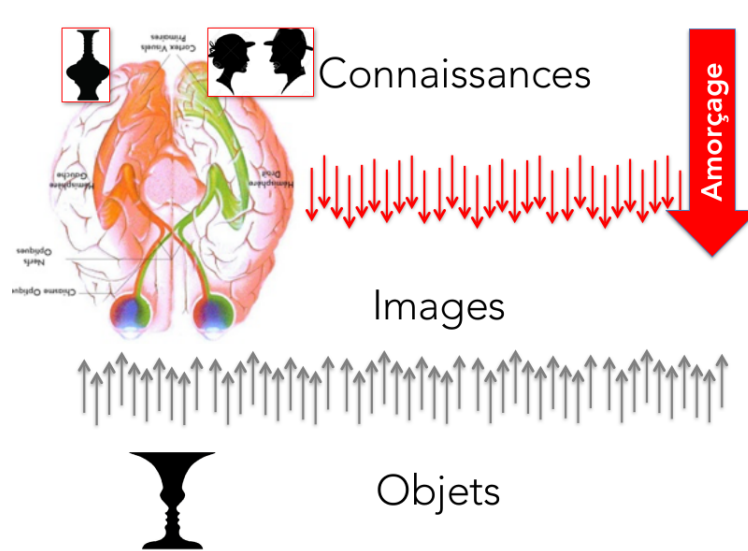
Lisbonne Berne Zurich Amsterdam Tokyo
 Geneva Manchester Rotterdam San Francisco
 Bruxelles Paris **Rome** New York Madrid
 Lisbonne Berne Zurich Amsterdam Tokyo
 Geneva Manchester Rotterdam San Francisco
 Bruxelles Paris Londres New York Madrid
 Lisbonne Berne Zurich Amsterdam Tokyo
 Geneva Manchester **Rotterdam** San Francisco
 Amsterdam Tokyo Geneva Manchester
 Rotterdam San Francisco Amsterdam Tokyo
 Geneva Manchester Rotterdam **Moscou**

Dans cette image, le taxi en couleur éclatante déclenche les processus bottom-up. Notre attention devrait donc surtout sélectionner le taxi. Mais un autre processus intervient: il n'est pas normal de voir un éléphant en liberté dans les rues d'une ville. Sa présence entre en conflit avec ce qu'on attend en regardant une image de ville nocturne... Dans ce cas, c'est notre cognition qui dirige notre attention et non pas les propriétés physiques de l'image qui guident l'attention. On parle dans ce cas de processus **top-down**. La perception d'objets dans une image est fortement influencée par nos attentes, nos expériences, nos connaissances,... La perception d'objets dans une image est fortement influencée par nos attentes, nos expériences, nos connaissances,...



L'effet d'amorçage ou priming effect consiste à présenter un stimulus (l'amorce) afin d'influencer la perception d'un autre stimulus (la cible).

L'amorce était les images préalables de vase ou de profils et la cible était l'image ambiguë. Nous percevons beaucoup d'objets mais les processus bottom-up guident notre attention sur certains objets particuliers, c'est l'attention sélective. Un exemple bien connu dans la perception auditive est le cocktail party effect: si je parle avec plusieurs personnes au milieu d'une soirée bruyante, je perçois néanmoins les conversations externes car, si quelqu'un prononce mon nom une autre table, ceci attirera néanmoins mon attention.



3.2.4 Perception de la profondeur

La perception de la profondeur est construite par le cerveau qui réconcilie les informations provenant des deux yeux. Chaque oeil envoie un demi-champ de vision à chaque hémisphère. Il utilise des indices différents:

1. Indices binoculaires: le principe de base est simple, l'angle de l'image sur la rétine détermine la distance d l'objet par triangulation.

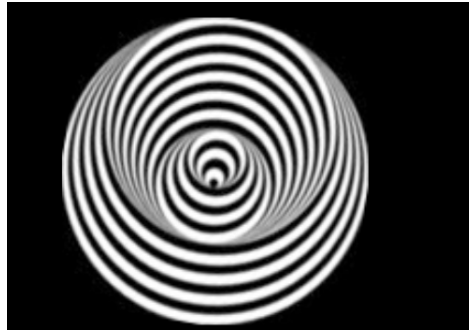
2. Indices monoculaires:

Même avec un seul oeil ouvert, nous avons une certaine perception de la profondeur grâce à différents indices.

La perspective: deux lignes parallèles qui s'éloignent du point de vision sont perçues comme se rapprochant. On peut en profiter avec un oeil fermé mais il nécessite un déplacement ce qui n'est pas très différent d'un principe binoculaire ou multioculaire.

L'effet parralaxe: c'est un méthode pour obtenir la profondeur en faisant déplacer les objets en premier plan plus rapidement que les objets en deuxième plan.

L'effet kinetic depth: les mouvements de l'objet donnent une perception de sa structure en 3D.



L’occlusion: même en 2D, cette image donne l’impression que le rectangle jaune se trouve en avant-plan du rectangle bleu. Cet effet peut être retourné, par exemple si la forme bleu n’est pas un rectangle mais un polygone en forme de U couché. Les ombres constituent un cas particulier d’occlusion.

La taille relative:

Malgré l’absence de lignes de perspective, les objets plus petits sont perçus comme plus éloignés, sauf si la sémantique de l’image nous fournit une autre interprétation de la différence de taille. Sans indices externes, on ne peut estimer la distance d’un objet dont on ne connaît pas la taille.

Position par rapport à l’horizon:

Le fait de briser la ligne d’horizon est une propriété des objets éloignés.

3.2.5 Recrer la profondeur en 2D

La différence de netteté de l’image entre l’avant plan et l’arrière plan fournit une perception de la profondeur.

Finesse de la texture:

La profondeur est communiquée en augmentant la finesse des textures en avant plan et en la réduisant dans le fond.

Brouillard de distance:

Un effet brouillard est utilisée dans les jeux pour donner une perception de la profondeur. Il permet aussi de ne pas calculer les décors très éloignés.

3.3 Remarques

Chaque hémisphère traite une demi-image et la partage avec l’autre hémisphère. Mais que se passe-t-il si l’information ne passe plus entre les deux hémisphères? Ceci arrive notamment dans des cas graves d’épilepsie qui conduit à une intervention chirurgical qui sectionne le corps calleux, le faisceau d’axones qui connecte les deux hémisphères.

Expérience du Split Brain: grâce à un dispositif technique, on projectte par exemple deux demi-images droites qui ne sont envoyées qu’à l’hémisphère gauche. Celui-ci étant spécialisé dans le traitement du langage, le sujet peut

dire "face". Si ces deux demi-images sont transmises à l'hémisphère droit, le sujet ne pourra pas lire le mot mais pourra dessiner l'objet.

Stimuli subliminaux:

Hypothèse controversée:

1. Si un stimulus visuel est inférieur au seuil de perception (en taille, durée, longueur d'ondes), il serait néanmoins perçu par notre inconscient. 2. Il aurait de ce fait une influence non-contrôlable sur notre comportement.

3.4 Exercises

Exercise 3.1.

Quelles composantes de l'oeil lui permettent de réaliser les performances suivantes?

- Lire des petits caractères à l'écran. (cristallin)
- Discriminer un mot en bleu d'un mot en rouge. (cônes)
- S'adapter à la distance de l'écran pour obtenir une image nette. (cristallin)
- Lisant du texte en haut de l'écran, détecter que l'icône mail se met à clignoter en bas d'écran (bâtonnets)
- Lors d'une promenade nocturne, reconnaître des objets dans la pénombre (bâtonnets)

Exercise 3.2.

Où se passent les phénomènes suivants?

- L'effet stroop (cerveau)
- La persistance de l'image (oeil)
- L'effet phi ou beta (cerveau)
- L'effet parallaxe (cerveau)