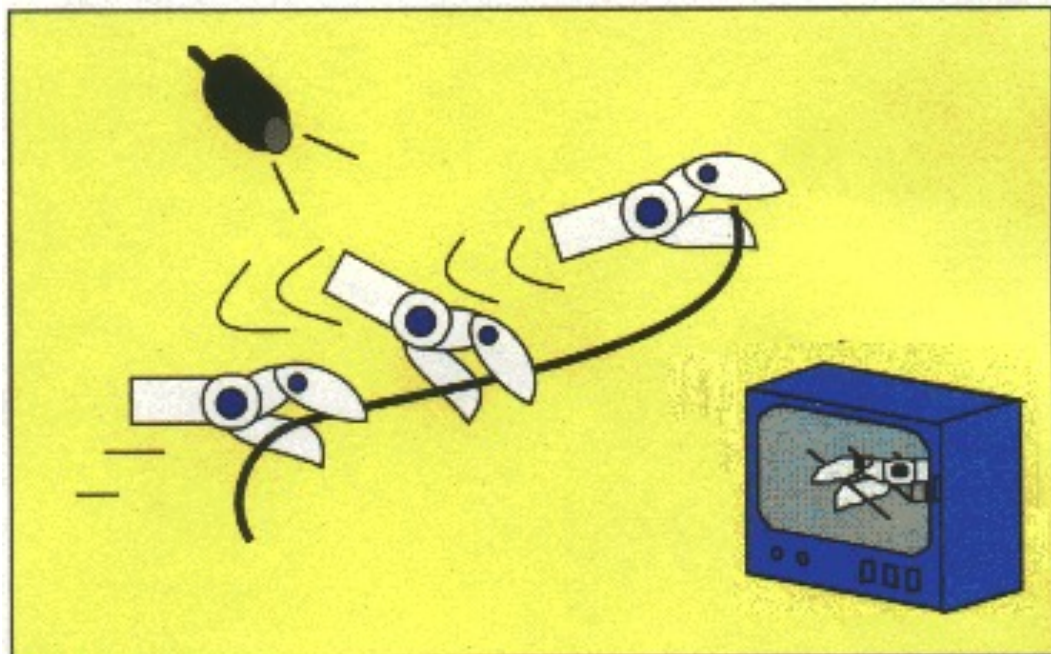


*Facultés de Technologies Industrielles (TIN),
et Formation En Emploi (FEE)
Filières microtechniques et électronique et automatisation industrielle (MI et EAI)
LaRA - Laboratoire de Robotique et Automatisation*

VISION PAR ORDINATEUR

Jean-Daniel Dessimoz



Yverdon-les-Bains, 17 février 2017

VISION PAR ORDINATEUR**Table des matières**

1	Introduction	1
1.1	Généralités	1
1.2	Histoire	5
1.3	Plan du chapitre	6
2	Acquisition d'images	6
2.1	De l'émission d'un rayon à sa représentation en mémoire	6
2.2	Sources de lumière	7
2.3	Objets	9
2.4	Milieu de transmission	10
2.5	Prismes anti-bougé	11
2.6	Filtres	11
2.7	Objectif	12
2.8	Capteurs	15
2.9	Échantillonnage	21
2.10	Quantification	23
2.11	Exemple de numérisation d'image	24
2.12	Composition des couleurs	25
2.13	Marche à suivre pour acquérir une image	31
3	Traitement bidimensionnel d'images	34
3.1	Opérations par pixels individuels (point par point)	35
3.2	Modifications des pixels d'après leur voisinage	37
3.3	Opérations globales ou par régions	48
3.4	Détection de mouvement	54
3.5	Transformée en cosinus	55
3.6	Temps de traitement	55
4	Analyse de scènes	57
4.1	Extraction de caractéristiques	58
4.2	Changement de représentation	62
4.3	Segmentation	65
4.4	Reconnaissance de "formes"	67
4.5	Estimation de distance	71
4.6	Problème général de l'estimation de pose	75
4.7	Inspection	76
4.8	Prise de pièces en vrac	78

5 Particularités de la vision pour robots	80
6 Matériel pour la vision	81
6.1 Systèmes standard	81
6.2 Carte de numérisation pour PC: Matrox	81
6.3 Traitement d'images par d'autres moyens	82
7 Exemples d'application	83
7.1 Tri de chocolats	83
7.2 Contrôle de circuits imprimés	83
7.3 Guidage visuel en soudure à l'arc.....	84
7.4 Manutention d'objets.....	84
7.5 ABB.....	85
7.6 Robots mobiles autonomes de la HEIG-VD	86
8. Conclusion	87
REFERENCES BIBLIOGRAPHIQUES	88
INDEX.....	89

VISION PAR ORDINATEUR

1 INTRODUCTION

Avant d'aborder le détail de techniques pour l'ingénieur, en vision par ordinateur, il convient de situer notre propos, de rappeler brièvement l'évolution historique de l'étude de la vision, et d'indiquer le plan suivi pour ce chapitre.

1.1 Généralités

Le domaine de la vision artificielle est vaste, aussi est-il nécessaire de l'aborder en plusieurs étapes. Mais avant de les présenter, cette section indique l'objectif, puis les limites que l'on s'est fixées.

1.1.1 Objectif

L'objectif du fascicule est de fournir à des ingénieurs les bases indispensables pour comprendre et donc utiliser, voire développer, les techniques de la vision par ordinateur. Ces techniques sont utiles dans de nombreux contextes. Citons-en quelques uns: contrôle de qualité, guidage de robots pour la maintenance et pour les procédés de fabrication, classification, ou analyse de clichés cellulaires en imagerie médicale.

1.1.2 Contexte de la vision par ordinateur

Le chapitre traite de la vision par ordinateur. Pour commencer, il est utile de situer cette discipline parmi d'autres. Nous le faisons d'abord en décrivant la place des senseurs d'images parmi les divers capteurs utilisés en automatisation industrielle. Puis nous tentons de distinguer les nombreuses composantes du traitement général des images.

Place des capteurs visuels parmi les autres

Les capteurs utilisés en automatisation sont de types variés. Ils perçoivent par exemple le contact, la proximité, les forces, les moments, les images et donc les formes et les distances. Le sens visuel repose sur l'utilisation de capteurs d'images. Ceux-ci ont en propre les caractéristiques suivantes:

- Très grand débit d'information. Un débit d'information de l'ordre de 100000 bit/s est courant pour les applications en vision artificielle. Ceci dépasse largement tous les autres capteurs utilisés en automatisation. Ainsi, les capteurs de contact ont un débit allant de 10 (interrupteur) à quelques centaines de bit/s (peau artificielle basée sur l'emploi de caoutchouc à résistance électrique dépendant de la pression ou de capteurs de pression intégrés). Les capteurs de forces et de moments délivrent de l'ordre de 1000 bit/s; cette information est généralement distribuée selon six composantes. D'autre part, plusieurs capteurs de proximité peuvent être montés sur la main des robots. L'amplitude des signaux de proximité reçus n'est généralement codée que de manière binaire (présence ou absence d'objets), car elle est influencé par de nombreux facteurs qu'il est difficile de démêler (réflectance, distance, angle des surfaces de réflexion . . .). La fréquence

maximale de ces signaux se confond avec celle des déplacements mécaniques. Les capteurs de proximité peuvent être capacitifs, inductifs, à ultrasons ou électromagnétiques (lumière visible ou infrarouge). Leur débit d'information est de l'ordre de 100 bit/s.

- Perception à distance de la géométrie d'une scène (indirectement, c'est-à-dire par interprétation de la réflectance des objets). Ceci est avantageux, car une optimisation des mouvements du robot peut en résulter. Par contraste, la plupart des autres capteurs nécessitent le déplacement du bras de robot pour explorer l'environnement, ce qui est très lent. Bien que les capteurs à ultrasons puissent aussi estimer la géométrie des objets à distance, ils sont peu utilisés car ils n'ont, dans l'air, qu'une résolution très grossière. Les ultrasons sont rapidement absorbés lorsque leur fréquence dépasse 400 kHz. La résolution est approximativement gouvernée par l'équation $d = \lambda/2$, où λ est la longueur d'onde de la porteuse. Les distances entre capteurs et objets sont estimées par différents moyens indirects. Une méthode parfois utilisée est basée sur la focalisation de l'image. Lorsque l'image est nette, la distance entre l'objet et le capteur est proportionnelle à la distance focale. Beaucoup plus utilisée, la triangulation se base sur la propriété que les six éléments d'un triangle (angles et côtés) sont fixés si un côté et deux angles sont connus. Le triangle est formé par deux capteurs et l'objet (vision stéréoscopique) ou par un capteur, l'objet et une source lumineuse directionnelle. Ce point est développé plus loin.
- Les limitations d'utilisation des capteurs visuels sont liés à l'impossibilité de percevoir les grandeurs dynamiques (forces et moments). Même les grandeurs géométriques ne sont pas toujours perceptibles. Par exemple, la réflectance d'une scène peut être trop uniforme ou, au contraire, trop variée pour permettre l'analyse. Des obstacles peuvent aussi se trouver entre capteur et objet. Typiquement, durant la phase terminale des mouvements du robot, la main cache une partie de la zone de travail.

Techniques diverses liées à l'image

Les capteurs visuels délivrent des images qui doivent être traitées. En traitement d'image, on peut distinguer deux catégories principales: la vision artificielle et l'infographie. La première catégorie traite de l'analyse et de la compréhension de scènes par des machines. La deuxième s'occupe de synthèse et de traitement d'images pour des observateurs humains. A ce deuxième type de technique correspond la majorité de l'équipement matériel existant. Les systèmes les plus performants traitent simultanément, en temps réel, plusieurs images de 4096x4096 points voire plus (p. ex. générées par satellites). Le traitement peut inclure des agrandissements, des rotations, des filtrages et des changements de couleurs.

En vision artificielle, de nombreuses applications se démarquent: inspection, classification, surveillance, contrôle de robots, etc. Pour la surveillance, des systèmes commerciaux détectent le mouvement; à l'Université de Karlsruhe, un système visuel surveille la zone de travail d'un robot pour prévenir les accidents. En classification, il existe par exemple des équipements de reconnaissance de caractères

pour aveugles, des systèmes pour tri postal ou des machines à compter les différentes sortes de globules blancs du sang. En inspection, le domaine des applications industrielles potentielles semble énorme. Exemples d'équipement très performant: des machines inspectent les soudures de circuits imprimés, à la cadence d'un circuit par seconde; d'autres contrôlent l'apparence de médicaments à des vitesses de plusieurs dizaines de capsules par seconde.

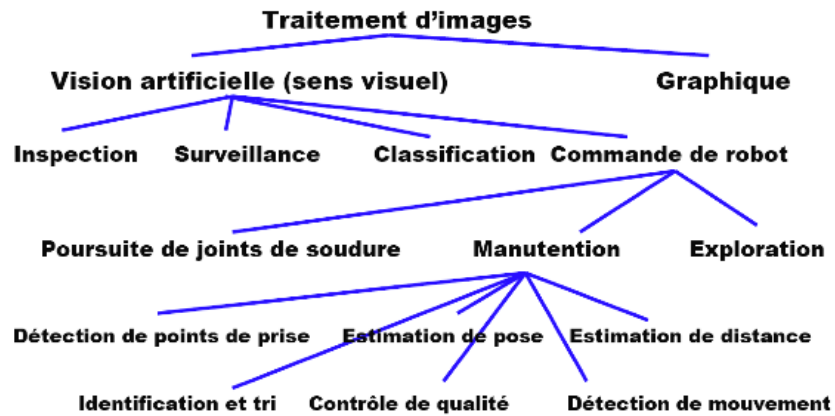


Fig. 1.1 La vision artificielle – contexte et quelques domaines particuliers

Spécifiquement pour le contrôle des robots, on peut mentionner les applications en soudure à l'arc, en exploration et en manutention. Pour la soudure, certaines firmes commercialisent des systèmes de poursuite de joints en deux passages (d'abord la reconnaissance visuelle, puis la soudure). Mais il existe également des systèmes capables d'observer le métal en fusion et de commander les déplacements de la torche en conséquence. Un sens visuel pour robots soudeurs est très demandé, car ces robots sont déjà installés en grand nombre dans les ateliers industriels. Pour l'exploration, le sens visuel est fondamental: les robots mobiles doivent optimiser leur trajectoire à travers des environnements souvent imprévisibles. Entre autres, le Jet Propulsion Laboratory aux USA et le LAAS au «Centre national de recherches scientifiques» (CNRS) de Toulouse consacrent d'importantes ressources à la recherche dans ce domaine. Le sens visuel doit être particulièrement performant en termes d'abstraction et d'intelligence pour ce type d'application.

Quant à la manutention d'objets, c'est le terrain idéal pour les robots industriels dotés de vision, en atelier. La perception de la géométrie de la scène est alors déterminante. Lorsque par contre le robot touche des éléments rigidement fixés (assemblage, usinage, etc.), les grandeurs dynamiques, de contact ou de proximité sont plus indiquées pour la commande de robot. Localisation de but, prise d'objets, estimation de pose (position et orientation), identification et tri de pièces sont des tâches typiquement dévolues au sens visuel durant la manutention. De plus, un contrôle de qualité, au moins grossier, est souvent nécessaire simultanément.

1.1.3 Vision artificielle: une chaîne à trois maillons

La vision artificielle est un domaine trop vaste et trop varié pour pouvoir s'approcher globalement. Il est au moins nécessaire de le partager en trois disciplines complémentaires: l'acquisition, le traitement bidimensionnel, et l'analyse d'images.

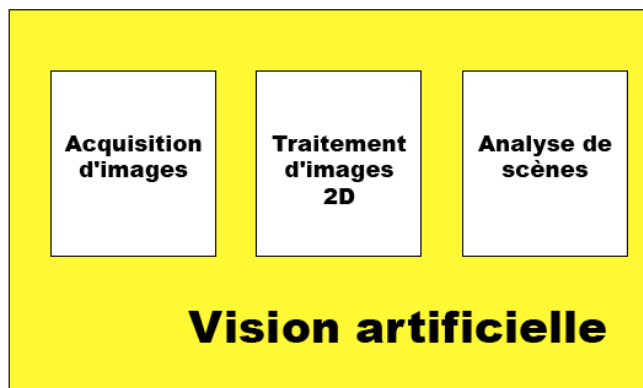


Fig. 1.1b La vision artificielle peut schématiquement se décomposer en trois domaines distincts

En pratique, les techniques d'acquisition d'images jouent un rôle beaucoup plus important qu'on ne le croit généralement. La prise d'image n'est jamais simple et conditionne pourtant de façon déterminante la suite du travail. Le problème est aggravé par le fait l'utilisateur est généralement peu formé aux techniques visuelles et qu'il est aussi le seul à connaître les détails de son application. Son apport reste critique.

Le traitement bidimensionnel d'images, c'est l'ensemble des opérations élémentaires qui à partir d'une structure de données bidimensionnelle (image) nous donnent un résultat de même type. Au contraire du précédent, le deuxième bloc de la vision artificielle ne pose généralement pas de problème pratique. Contrairement à l'opinion répandue, ce bloc est assez "modeste" et reste presque négligeable par rapport aux deux autres. La limite principale est posée ici par le grand nombre d'éléments qu'il faut éventuellement traiter. Mais il y a peu de traitements judicieux à faire à ce niveau, et lorsque c'est le cas, les opérations sont simples, régulières, et peuvent être assumées par des circuits intégrés spécialisés.

Enfin, l'image acquise et (pré-)traitée doit être analysée. D'une complexité virtuellement sans limite, cette tâche est encore assez simple aujourd'hui pour la plupart des applications. Ce qui frappe, ce n'est pas tant la complexité d'un quelconque algorithme considéré individuellement, mais plutôt l'extraordinaire diversité qui, ensemble, les caractérise.

Domaine	Limites actuelles majeures	Facteurs d'amélioration	Importance relative
Acquisition d'images	Diversité des applications	Savoir-faire de l'utilisateur	50%
Traitement 2D	Grand nombre d'opérations	Circuits spécialisés, traitement sélectif	10%
Analyse de scène	Complexité des algorithmes	Meilleurs algorithmes, processeurs plus rapides	50%

Fig. 1.2 Dans les trois domaines de la vision, des problèmes spécifiques doivent être résolus

1.2 Histoire

De tous temps, des chercheurs renommés ont essayé de comprendre le phénomène de la vision. Les premières études connues remontent probablement à Euclide, qui a essayé au 3ème siècle avant J.-C. d'en formaliser certaines lois fondamentales. Deux des lois d'Euclide sont par exemple *Les lignes droites issues de l'œil franchissent des distances d'une grande longueur*, et *Ce que l'on voit sous des angles plus nombreux apparaît plus net*; Euclide a formulé sept lois de ce type. Plus près de nous, au siècle passé, deux grands noms de la science ont apporté leur contribution au problème: von Helmholtz et Mach. A partir des années 1920, les psychologues de la fameuse école de la Gestalt se sont également penchés sur les mécanismes de la perception visuelle.

La formalisation dès 1930 des bases théoriques du traitement des signaux (N. Wiener, C. Shannon, etc.) a conduit au démarrage des approches analytiques. Le traitement numérique d'images a fait son apparition dès les années 1950, en parallèle avec le développement de l'informatique. Le premier grand essor des méthodes que l'on pourrait qualifier de "classiques" date de la fin des années 1960.

Un deuxième coup de fouet essentiel est venu de D. Marr [Mar82], qui dès 1975 a jeté les fondements de ce que l'on appelle maintenant la "computational vision". Cette expression est difficile à traduire, car elle implique beaucoup de sous-entendus. Par opposition aux approches classiques, on cherche ici à construire une théorie générale qui devrait éviter d'avoir à traiter les problèmes de manière ad-hoc. Les premiers liens avec la neurophysiologie, l'intelligence artificielle, la psychologie cognitive, etc. datent principalement de cette époque.

Dès le début des années 1980, le mouvement général de la recherche en traitement et analyse d'images se dirigeait vers la vision artificielle.

1.3 Plan du chapitre

Le chapitre est organisé de la façon suivante. Les trois maillons principaux de la vision artificielle que nous venons d'introduire, l'acquisition, le traitement 2D, et l'analyse font tour à tour l'objet des chapitres 2, 3 et 4. Des éclairages complémentaires sont ensuite successivement dirigés sur le cas particulier de l'utilisation conjointe de robots, sur les divers types de matériel électronique assurant la vision artificielle, ainsi que sur quelques exemples représentatifs de ce qui se fait actuellement.

2 ACQUISITION D'IMAGES

Une image est essentiellement la projection d'une scène lumineuse tridimensionnelle sur un plan. Il y a donc, en principe, perte d'information.

Dans une certaine mesure, en multipliant les projections selon des angles différents, il est possible de connaître la scène dans ses trois dimensions. Cette approche est utilisée en vision stéréoscopique par exemple.

Une autre manière, indirecte, de connaître la troisième dimension consiste à projeter un éclairage latéral sur la scène.

Acquérir une image, c'est transférer une information visuelle émanant du monde réel jusque dans la mémoire d'un ordinateur, via un grand nombre d'étapes. Après une brève vue d'ensemble, nous allons nous arrêter à chacune de ces étapes, et en discuter les paramètres les plus significatifs. Bien qu'on les examine ici individuellement, il est clair que ces paramètres sont souvent liés entre eux. Par exemple, on peut souvent remplacer une lumière colorée par un filtre sur l'objectif, ou encore, une quantification à niveaux de gris multiples n'a guère de sens en diascopie.

Ce paragraphe se termine avec quelques considérations sur la (dé-)composition des couleurs.

2.1 De l'émission d'un rayon à sa représentation en mémoire

Représenter en mémoire le monde réel en vue d'analyses ultérieures, c'est mettre en oeuvre une chaîne complexe schématiquement illustrée sur la figure ci-dessous.

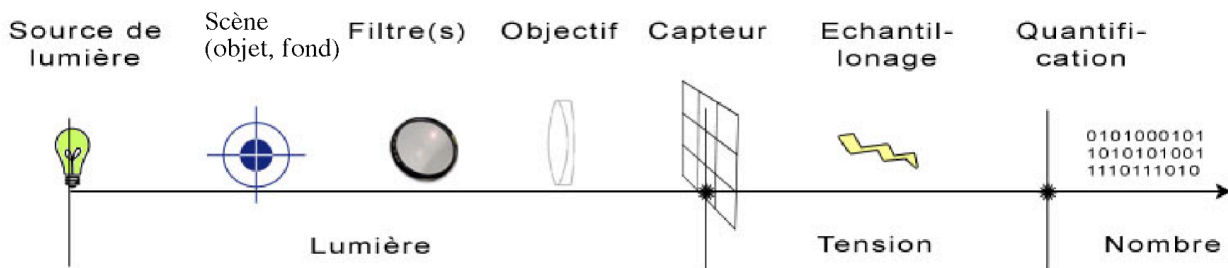


Fig. 2.1 Chaîne d'acquisition de l'information visuelle (fig. C. Nicolas 2005).

On pourrait y ajouter un maillon, entre objet et objectif, afin de tenir compte de l'influence du milieu de transmission. Celui-ci est généralement parfaitement transparent, mais il y a parfois d'autres cas, tels

celui de l'observation par avion ou par satellites, ou encore lorsqu'on utilise des torches de fibres optiques.

Le but de l'acquisition d'image, c'est, *semble-t-il*, de représenter fidèlement en mémoire une scène, c'est-à-dire la portion du monde se trouvant devant la caméra.

En fait le but est différent: il s'agit de saisir avec une intensité maximale les éléments particuliers significatifs pour une application donnée, et de "gommer" le mieux possible toute autre information.

2.2 Sources de lumière

Il est parfois possible de travailler en lumière ambiante. Mais cela est rarement le cas, car les perturbations possibles sont multiples: rayons de soleil, ombres portées par le personnel, etc. De plus, l'application requiert souvent que l'on mette en évidence certains détails par un éclairage particulier.

Caractéristiques spectrales (blanc, couleur, visible, monochrome ...)

Voyons d'abord les situations courantes. La lumière naturelle est à peu près blanche. L'éclairage artificiel à filaments offre relativement peu de bleu, avec au contraire beaucoup d'infrarouges. Les sources fluorescentes ("néon") sont, à part quelques variantes, plus généreuses en bleu, et n'émettent pratiquement pas d'infrarouges.

Il est parfois intéressant d'utiliser un éclairage à laser. C'est toujours à cause de sa lumière monochromatique qu'on le fait. Ceci permet d'une part de focaliser de façon très précise le rayon lumineux. D'autre part, puisque toute la puissance est concentrée sur une seule longueur d'onde, cela permet de travailler dans un très petit domaine spectral, et de surpasser ainsi d'autres signaux lumineux. Quelques fois, c'est en plus à de la lumière cohérente que l'on recourt parfois, typiquement pour détecter par interférométrie de petites distances.

Notons que les techniques de traitement d'images s'appliquent aussi à des signaux non-lumineux, tel qu'un rayonnement gamma, ou un balayage à ultrasons.

Structure

Il est en général souhaitable d'avoir un éclairage uniforme sur la scène. Mais parfois, une structure particulière est utile.

Dans certaines applications, c'est un point lumineux unique qui est émis (par ex. pour la mesure de distance). D'autres fois, pour gagner du temps, on tente d'émettre plusieurs points, en réseau, ou en ligne. Le problème qui peut alors surgir, c'est l'ambiguïté entre les points perçus: à quel point émis chacun des points perçus correspond-il?

Dans le cas de l'émission de lignes de lumières, on parle aussi parfois de "plans lumineux" pour des raisons évidentes.

Modulation

Contrairement aux apparences, l'éclairage est rarement d'intensité constante. Même dans le cas de la lumière naturelle, il en va pratiquement jamais ainsi: L'utilisateur d'un système à gain fixe est généralement surpris de constater combien le passage aux conditions extrêmes - scène obscure ou saturation- est à la fois aléatoire et rapide.

En éclairage artificiel, une modulation liée à la fréquence du réseau - 50 Hz en Europe - est très généralement présente. Bien que moins marquée pour les lampes à filaments que pour les lampes fluorescentes, elles ne sont nullement négligeables. Il se trouve que la fréquence standard de balayage pour les caméras est également de 50 Hz, d'où une "annulation" du phénomène (l'intégration par le capteur de l'intensité lumineuse incidente sur 1/50 s correspond à un filtrage de toutes les harmoniques de 50 Hz).

Lorsque la fréquence de lecture du capteur n'est pas égale à 50Hz (ou l'un de ses sous-multiples), comme c'est fréquemment le cas pour les capteurs linéaires, des mesures spéciales, telles que l'éclairage sur source continue doivent être envisagées.

Une autre modulation de la source est aussi utile dans certains cas. Par exemple pour les pièces en mouvement (cycliques ou non), il est possible d'éviter un "bougé", c'est-à-dire d'obtenir une image nette, en utilisant un stroboscope. La durée d'un éclair est de l'ordre de quelques μ s. Une scène en mouvement peut aussi être saisie à l'aide d'un obturateur rapide (quelques ms), mais il faut alors un éclairage continu intense; par contre on a plus de liberté dans le choix des caractéristiques spectrales de la lumière.

Lorsqu'un seul point est émis, on peut également agir sur son intensité afin de "voir" les zones sombres, sans toutefois éblouir le capteur au passage de surfaces claires ou réfléchissantes.

Position de la source

Il convient tout d'abord de distinguer entre épiscopie et diascopie. Le premier cas est le plus courant. C'est celui où l'on éclaire la scène "par-dessus". La lumière est réfléchiée vers le capteur. En diascopie, c'est la lumière directement transmise par la source que l'on capte. La scène se distingue alors par une absorption plus ou moins grande. Souvent, c'est alors simplement la silhouette opaque de l'objet que l'on désire percevoir.

On parle ci-dessus d'éclairage homogène ou diffus. Mais parfois il vaut mieux utiliser une source ponctuelle ou dirigée. Suivant la position relative de la source, d'une surface réfléchissante, et enfin du capteur, l'intensité finalement perçue varie. De cas en cas, une surface brillante (réflexion spéculaire) va renvoyer beaucoup plus de lumière qu'une surface mate. Pour d'autres orientations, nous aurons l'effet inverse. L'utilisation systématique de sources ponctuelles situées à différents endroits permet parfois de retrouver la forme tridimensionnelle des objets, à partir des normales respectives des différents éléments de surface de la scène.

Enfin dans le cas de l'estimation de distance que nous discuterons plus loin, la direction de l'éclairage joue un rôle déterminant.

2.3 Objets

Que l'on veuille faire ressortir un objet (ou un élément d'objet) dans la scène, ou qu'on veuille au contraire le "camoufler", il est nécessaire d'exploiter l'une ou l'autre de ses caractéristiques affectant la lumière.

Réflectance

La première caractéristique, c'est la réflectance. Cette propriété est caractérisée par un coefficient liant l'énergie réfléchie à celle de la lumière incidente. Ce coefficient est proche de 1 pour un objet "blanc", et voisin de 0 pour un "noir". Entre deux toute une gamme de "gris" est possible.

Il arrive également qu'une partie de la lumière ne soit pas réfléchie mais au contraire traverse l'objet (pièce translucide).

État de surface

Les données essentielles sont ici la nature brillante ou au contraire mate des objets. Dans le premier cas, la lumière est réfléchie selon une direction préférentielle, à la manière des miroir. De plus la lumière réfléchie est polarisée dans un plan perpendiculaire à la surface de réflexion.

On considère généralement que dans le cas d'une surface mate, l'intensité d'un rayon de lumière réfléchie varie comme le cosinus de l'angle défini par lui et une normale à la surface (loi de Lambert).

Entre les cas extrêmes que nous avons discutés ci-dessus, toutes sortes d'autres caractéristiques superficielles sont possibles, telles par exemple l'aspect "pelure d'orange" de certains revêtements, ou la rugosité des pièces de fonte, qui posent des problèmes peu prévisibles en acquisition d'images.

Couleur

La réflectance définie ci-dessus peut dépendre du contenu spectral de la lumière incidente. L'objet éclairé en lumière blanche apparaît alors coloré. Suivant les cas, il peut être judicieux d'exploiter telle fenêtre spectrale ou telle autre. Le choix se fait par l'utilisation d'un éclairage, d'un filtre, ou d'un capteur appropriés.

Mouvement

Il est important de savoir si la scène bouge ou non, car les temps d'exposition doivent être adaptés en conséquence. Par exemple si l'objet en mouvement gêne, on peut tenter de prendre la scène avec en temps d'exposition relativement grand, et une ouverture réduite en conséquence. A l'inverse, s'il s'agit d'un élément intéressant, un temps d'ouverture ou d'éclairage bref seront à considérer.

Orientation

Le résultat de l'acquisition d'images dépend généralement de l'orientation relative de la source de lumière, des surfaces visibles, et du capteur. Ce point a été discuté au § Source de lumière.

Complexité

La surface visible des objets pose des problèmes de nature diverse.

Dans les cas les plus simples, la surface peut être simplement caractérisée par un indice uniforme de réflectance (le niveau de gris) ou par le fait que l'objet soit opaque (cas de la diascopie).

Une situation un plus difficile est celle où la surface se distingue par des caractéristiques particulières de couleur, ou par des détails de nature structurels (par exemple une cheminée au milieu d'un toit, ou une vis sur un couvercle).

Dans le cas général où l'indice de réflectance n'est pas uniforme, mais que ses variations sont homogènes, on parle de texture. Il peut s'agir par exemple d'une surface sablée, d'une coupe dans le bois, ou d'un tissu écossais. La notion d'homogénéité apparaît parfois à un niveau d'abstraction très élevé. Pensons par exemple à un motif de type "personnages de Walt Disney" sur un papier peint.

De façon générale, dans toutes les situations où les détails significatifs sont du même ordre de grandeur que les motifs élémentaires de la texture ("texel"), la vision artificielle est aujourd'hui fortement limitée.

2.4 Milieu de transmission

De façon générale, la lumière voyage à travers un certain milieu pour aller de l'objet au capteur. Il peut s'agir de l'air (turbulences, dans le cas des grandes distances ou des grands différentiels de température), de fumée, ou d'autres conditions encore.

En particulier, on utilise parfois des torches de fibres optiques, où l'ordre entre voisin est maintenu entre face de départ et face d'arrivée.

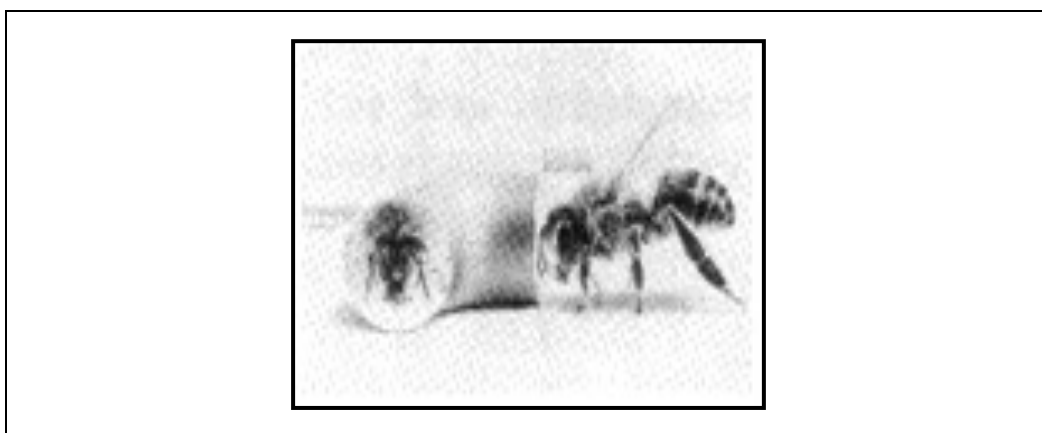


Fig. 2.1b Transmission d'image à travers une torche de 7000 fibres

2.5 Prismes anti-bougé

Sur certains caméscopes, les dimensions ont beaucoup diminué, ce qui entraîne en conséquence une diminution sensible d'inertie et donc un risque accru de bougé. Pour faire face à ce problème, certains fabricants ont intercalé un prisme sur le chemin lumineux, et ils le font varier légèrement d'orientation en temps réel de façon à limiter l'effet désagréable de bougé. D'autres fabricants opèrent la même correction à l'aide de méthodes "numériques" sur la base d'une image brute non directement accessible à l'utilisateur.

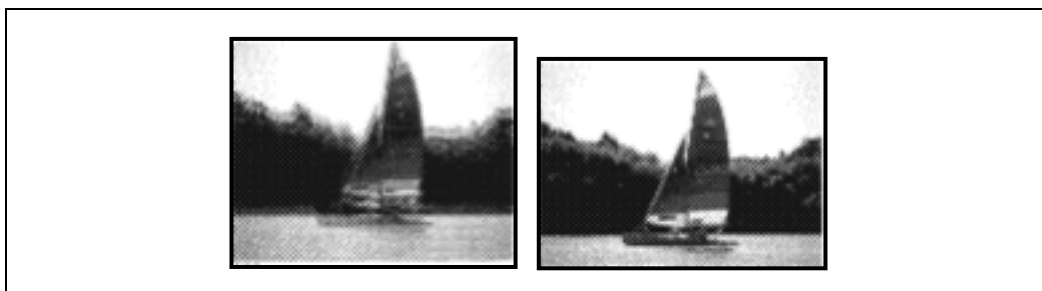


Fig. 2.1c Compensation de bougé par traitement numérique temps réel

2.6 Filtres

Les filtres, dans le langage courant, ont pour propriété de retenir certains éléments et d'en laisser passer certains autres. Ici, il s'agit de laisser passer le signal visuel, tout en éliminant certaines de ses composantes. Le filtre peut agir sur le spectre de lumière (couleurs, infrarouge...) mais aussi travailler selon d'autres paramètres, tel que la fréquence *spatiale*, ou le plan de polarisation.

Couleurs

Les filtres couleurs sont très courants en photographie. Ils sont disponibles en grande variété, pour un coût modeste. Ils peuvent en général s'utiliser valablement tant au niveau de la source de lumière, qu'à celui du capteur, quoique celui-ci soit un peu plus favorable.

Pour des caractéristiques extrêmement sélectives, telles qu'on en désire parfois lorsqu'on utilise des lasers, il faut recourir aux filtres interférométriques.

Infrarouge

Suivant les applications, il peut être utile d'éliminer la portion visible du signal lumineux, ou au contraire ne retenir qu'elle. Dans le premier cas, des filtres gélatines sont très efficaces. S'il faut au contraire absorber les infrarouges, des filtres plus lourds, épais et fragiles sont nécessaires.

Passe-bas spatiale

Les variations de l'indice de réflectance, sur une surface, peuvent se succéder avec un certain nombre de cycles par mm. On parle alors de fréquences spatiales. La plupart des filtres spatiaux sont réalisés sous forme d'algorithmes ou de circuits électroniques spécialisés, et agissent séquentiellement sur chaque portion d'images.

Mais dans le cas du filtre passe-bas, une alternative rapide et peu coûteuse est fournie par l'objectif. Moins la mise au point est nette, et plus le filtrage spatial s'opère à fréquence de coupure basse. L'estimation exacte de la fréquence de coupure repose sur l'examen du comportement des rayons lumineux selon les lois de l'optique géométrique.

Polarisation

La lumière est une onde électromagnétique. Celle-ci se caractérise par des plans préférentiels pour ses composantes électriques et magnétiques. L'orientation de ces plans joue un rôle important dans bien des applications. Il s'agit alors de les filtrer (élimination de plans de polarisation selon des angles indésirables par un filtre "analyseur") en fonction d'une polarisation qu'on leur a imposée au préalable (filtre "polariseur") ou pour prendre en compte certains phénomènes de réflexion. Les qualificatifs polariseur et analyseur correspondent à la fonction désirée et non pas au filtre lui-même, qui peut avoir les mêmes caractéristiques intrinsèques.

2.7 Objectif

L'objectif doit "concentrer" le flux lumineux incident sur une caméra, et le focaliser sur l'élément transducteur (puce, film, tube...). Ses caractéristiques principales sont bien connues des photographes amateurs.

Focale

La focale est la propriété fondamentale d'un objectif. Elle indique la taille relative de l'image sur le capteur par rapport aux caractéristiques géométriques de l'objet (grandeur et distance à l'objectif). L'équation fondamentale est ici la suivante:

$$\frac{1}{f} = \frac{1}{f'} + \frac{1}{f''}$$

où f' et f'' sont les distances séparant objectif (dans le cas le plus simple: la lentille) de l'objet et du plan où l'image est nette (plan focal). Le facteur d'échelle est donné par le rapport de f'' à f' .

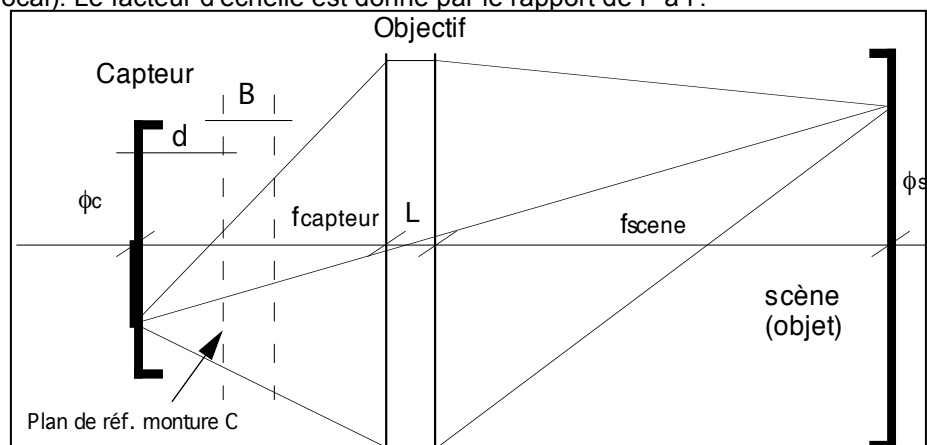


Fig. 2.1.d La course de réglage de la distance fcapteur peut simplement s'augmenter avec des tubes supplémentaires

Certains objectifs permettent de varier la focale ("zoom"). On peut aussi faire simplement un grossissement en rapprochant la caméra de l'objet et par addition, si nécessaire en conséquence, de tubes entre objectif et capteur.

(Distance, bagues)=f(focale, largeur champ de vue, « diamètre » capteur)

Données

focale (ex. :50mm)	30	30	30
largeur champ de vue [mm] :	10	10	100
« diamètre » capteur [pouce] :	0.67	0.67	0.67

Calculs

pouce [mm] :	25.4		
largeur capteur [mm]			
=0.7*0.8* »diamètre «	10	9	9
grossissement=larg. Capt/larg. Champ de vue			
grossissement [x] :	0.95	0.948266667	0.094826667
d [mm] =(estimé)	12	12	12
fcapteur=grossissement* fscene			
1/f=1/fscene+1/fcapteur			
fscene=(grossissement+1)/grossissement*f			
fscene=	61	62	346
fcapteur=	59	58	33

Résultats

Distance [mm] :	61	61.63667042	346.3667042
bagues [mm] :	17	16.448	0

Tableau de calcul de la distance à prévoir entre caméra et scènes, et grandeurs des éventuelles bagues additionnelles, en fonction de la focale de l'objectif, de la taille du champ de vue et de la taille du capteur d'image

Fig. 2.1.e Avec un objectif "œil de poisson", des distorsions semblent inévitables, mais les variations locales d'échelle sont en fait dues à la perspective



Plutôt que de focale, on parle parfois en angle d'ouverture. Une petite focale correspond à un grand angle.

Ouverture

L'ouverture, O, caractérise la quantité de flux lumineux qu'un objectif peut recevoir. Alors qu'il serait logique, semble-t-il, de donner la surface du trou par où la lumière passe, l'usage consiste à exprimer cette grandeur par un nombre, fonction du diamètre de ce trou, d, et de la focale, f:

$$O = \frac{f}{d}$$

L'objectif se caractérise par l'ouverture maximale. En cours d'utilisation, l'ouverture peut varier grâce à un diaphragme.

Profondeur de champ

Le plan focal est généralement ... plan. Dans ces conditions, un objectif ne donne une image nette que pour une distance déterminée. Tout point-objet se trouvant à une distance différente donnera non pas un point-image net, mais plutôt une tache de diamètre variable. Il apparaîtra flou.

En pratique, la résolution de la cible est toujours limitée ("grain" du film, densité de photodiodes, etc.). En conséquence, il n'est pas nécessaire d'avoir un point-image parfaitement net. Un certain diamètre sur la "tache"-image est admissible. Cela veut dire que l'on pourra avoir une image admise comme nette non seulement à la distance f , mais également un peu devant et un peu plus loin. La différence entre ces deux distances extrêmes, c'est la profondeur de champ.

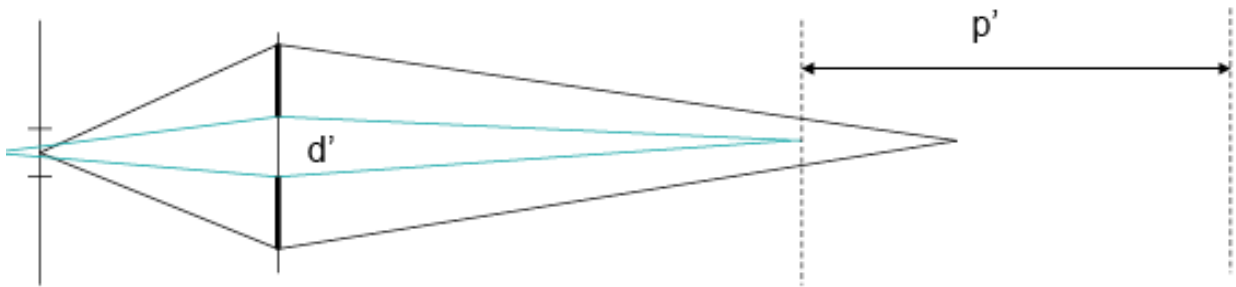


Fig. 2.1.f Avec une petite ouverture, le flou lié à une mise-au-point inexacte est d'importance réduite. On a une tolérance plus grande sur l'exactitude requise pour la distance objet-caméra (profondeur de champ, p')

La profondeur de champ augmente lorsqu'on réduit l'ouverture (fermeture du diaphragme) et lorsque la focale diminue.

Mise au point

La mise au point concerne l'ajustement de f , tel que défini plus haut (§focale), de façon à ce que chaque point de la scène donne un point également - et non pas une tache- sur le capteur.

Une mauvaise mise au point provoque un flou, et ceci est effectivement un filtrage passe-bas de l'image réalisé optiquement. Le phénomène est exploité dans les appareils automatiques: par analyse d'image dans des zones d'intérêt, le système peut juger de l'amélioration en fonction de variation de f , et s'efforce alors de converger vers une valeur optimale.

Distorsions géométriques

Les objectifs ne sont jamais parfaits. En particulier, les proportions d'un objet ne se retrouvent jamais exactement sur l'image. On pourrait aussi interpréter le "plan" focal d'un objectif comme une surface

bosselée. En pratique, on considère les distorsions géométriques dues à l'objectif comme étant généralement limitées à 1 %. Les distorsions géométriques sont également discutées au § 2.6 Capteurs.

Aberrations chromatiques

Chaque composante spectrale de la lumière (chaque longueur d'onde) suit un trajet légèrement différent dans l'objectif. Ce phénomène est exploité lorsqu'on veut précisément en étudier le spectre. En acquisition d'images, le phénomène est par contre plutôt gênant, conduisant aux erreurs d'aberrations chromatiques. Dans le visible, les objectifs sont en général assez bien construits pour éviter tout problème pratique. Mais il faut tenir compte du phénomène lorsqu'on en déborde (par ex. infrarouges), ou que des jeux spéciaux de lentilles sont considérés.

Autres remarques

Des lentilles et objectifs spéciaux permettent une foule de variantes : lentilles cylindriques, asphériques, télécentriques, « omnidirectionnelles », etc. , Par ailleurs les objectifs se montent généralement avec la norme « C » (filetage de diamètre adapté pour le 16mm) ou sa variante 5mm plus compacte, « CS ».

2.8 Capteurs

Le sens visuel repose sur l'utilisation de capteurs qui se classent en deux catégories: caméras à tube et caméras à semi-conducteur. Voyons d'abord globalement les avantages respectifs de ces deux techniques. Ensuite, chacun des termes les plus significatifs dans ce domaine seront définis.

Caméras à tube

Les caméras à tube sont les éléments traditionnels en prise d'images. Les tubes existent en grande variété. Ils ont en commun une bonne résolution spatiale. On peut typiquement distinguer jusqu'à 300 lignes noires sur fond blanc dans une image. Ainsi, on acquiert aisément 512 x 512 points significatifs par image (c'est-à-dire 512 lignes et 512 colonnes). Ces caméras ont aussi l'avantage d'être compatibles avec la majorité des équipements vidéo grand public (moniteurs, etc.). Une abondance de tubes à caractéristiques de transfert spécifiques sont disponibles. Les défauts sont principalement dans les distorsions géométriques de l'image (typiquement 2%) et dans la cadence d'acquisition. Celle-ci est généralement de 25 images par seconde pour la pleine résolution et de 50 images par seconde si l'on supprime l'entrelaçage.

Caméras à semi-conducteurs

La seconde catégorie de capteurs d'images regroupe les caméras à semi-conducteur. Lorsqu'elles contiennent un capteur bidimensionnel, elles ont une résolution généralement moindre que les caméras à tube, mais l'écart entre eux diminue et la gamme de variation est très grande. Il existe maintenant des capteurs à semi-conducteurs très petits (1/7 pouce, typique pour téléphone mobile) ou au contraire comprenant jusqu'à 4500x3000 pixels! (cf. par ex. caméra Kodak DCS Pro 14n, de taille 24x36mm)

En version industrielle, l'image perçue est souvent noir-blanc. Peut-être parce que le marché est trop petit, ces caméras sont chères. Une partie de l'explication se trouve aussi peut-être dans la qualité c'est-à-dire le nombre faible de cellules défectueuses.

Dans leur version grand public, les capteurs à semi-conducteur sont remarquablement bon marché. Ils ont maintenant totalement submergé le marché des appareils de photo et de la vidéo pour amateur. Le nombre de cellules est assez grand (plus de 10^6 pixels). Chacune est recouverte d'un filtre rouge, vert ou bleu, ce qui simplifie énormément la réalisation opto-mécanique.



Fig. 2.1.g Panasonic NV-S7. Structure interne avec lentille asphérique, filtre infra-rouge, et capteur 1/3 de pouce. Et Sony Exmor, avec couche sensible en surface (aire max).

Beaucoup de capteurs à semi-conducteur sont des capteurs linéaires. Les plus denses comptent 12'000 cellules et plus (p. ex. Atmel). Ces capteurs sont particulièrement intéressants lorsque les objets bougent, car ceux-ci assurent le balayage dans une direction et, de plus, le temps d'intégration court propre aux capteurs linéaires est alors un avantage.

Pratiquement exemptes de distorsions géométriques, les caméras à semi-conducteurs s'avèrent aussi plus sensibles, légères et compactes, que les caméras à tube. De plus, elles sont moins sujettes à l'éblouissement («blooming») et se satisfont de tensions d'alimentation basses. Par contre leur plage de sensibilité spectrale est limitée par l'usage exclusif du silicium.

Fig. 2.1.h Caméra omnidirectionnelle (<http://www.idmind.pt/english/research/hr.php>). Comme alternative, un réflecteur tout simple est fourni par une ampoule pour spot, dont le fond en "demi-sphère" est doublé en aluminium brillant.



CAPTEUR	TUBE	SEMI-COND.
Fonction de transfert (spectre de sensib., linéarité)	variable	fixe:silicium
Mode de lecture (norme TV, entrelaçage)	fixe:TV	variable
Résolution	++	variable
Distorsions géométriques	≤ 2 %	++
Effet de Moiré (recouvrement de spectre)	+	variable
Gain automatique	variable	variable
Risque de destruction (surexposition)	variable	++
Temps d'intégration, rémanence, blooming	variable	variable

Fig. 2.2 Caractéristiques majeures des caméras à tube et à semi-conducteur.

Principes physiques de transduction

Dans les tubes et les semi-conducteurs, il y a déplacement de charges électriques sous l'effet des photons incidents. Le courant ainsi provoqué (ou celui de rétablissement, qui remet le système à l'équilibre) sert de signal de mesure.

Sur le papier, ou dans les films, il y a réaction chimique par excitation lumineuse, d'où changement de couleur ou de solubilité.

En impression, divers principes sont utilisés: dépose d'encre, transfert de particules noires ou colorées par déplacement de charges et forces électrostatiques résultantes.

Voir aussi pp. 12-25 et 33-36 dans M. Kunt, G. Granlund et M. Kocher, " Traitement numérique des images", Presses Polytechniques Romandes, 1993, pp. 562 .

Fonction de transfert

La fonction de transfert indique l'intensité du signal perçu en fonction de la longueur d'onde (spectre de couleur) de la lumière incidente. Alors que pour les tubes une grande variété de substances chimiques sont disponibles, pour les capteurs à semi-conducteur, c'est le silicium qui est pratiquement toujours utilisé. La plage de sensibilité est alors pour moitié dans l'infrarouge, avec un pic dans le rouge, et une sensibilité très faible (1% du maximum) dans le bleu. Dans le visible, les couleurs peuvent à peu près être équilibrées par des filtres appropriés, mais ceci se fait au détriment de la sensibilité.

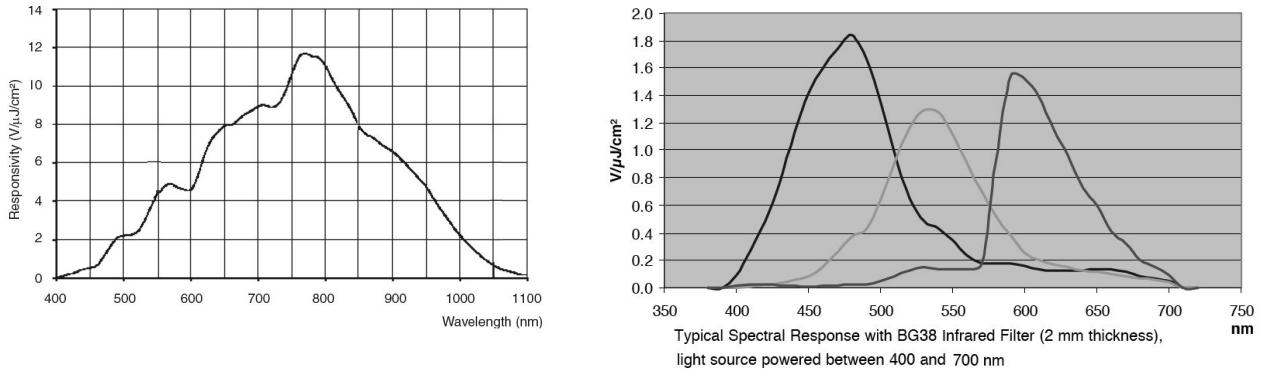


Fig.2,2b Fonctions de transfert typiques pour caméras à semiconducteurs (docs Atmel). A gauche, le capteur est brut, plus sensible dans l'infrarouge que dans le visible. A droite des mesures permettent de gérer les couleurs : filtres RVB intégrés à la caméra et filtre bloquant les infrarouges en supplément.

Mode de lecture

En général, c'est le balayage généralement adopté en télévision qui est utilisé tant pour les caméras à tube que pour celles, bidimensionnelles, à semi-conducteur. Il existe cependant des applications où des modes différents de lecture sont adoptés. C'est en particulier le cas lorsqu'on désire une très grande résolution, sur les caméras à tube, ou une très grande vitesse de prise de vue.

Résolution spatiale

La résolution correspond aux plus petits détails que l'on peut percevoir. La résolution du capteur n'a guère de sens, considérée isolément, car avec une optique adéquate, on peut y faire correspondre des détails de taille largement variables, sur la scène. Il faut en général la considérer conjointement avec la grandeur du champ de vue.

En pratique pour décrire la résolution, on se contente d'indiquer le nombre de lignes distinguables par mm (cas des tubes), ou le nombre de cellules sensibles (cas des semi-conducteurs).

Si une très grande résolution est nécessaire, le choix se porte sur un ou plusieurs capteurs linéaires. Pour la deuxième dimension, un balayage est réalisé par miroirs ou par un déplacement physique du capteur relativement à l'objet vu.

Ou il arrive aussi que l'on fasse plusieurs acquisitions successives de la scène, avec un léger décalage optique. Ainsi, par entrelaçage, une caméra Kodak au départ comprenant un capteur de 1'400*1'000

pixels avec pixels distants de $6.8 \mu\text{m}$ permet, chez un industriel romand de multiplier encore par 4 la résolution effective ("superrésolution"). Le décalage est assuré par une légère rotation d'un filtre infrarouge (de tels filtres sont généralement présents dans les caméras à semi-conducteur).

En 2011, le site Wikipedia indique un maximum de résolution pour un capteur photographique comme suit :

Hauteur	Largeur	Format	Nombre de pixels	Soit en Megapixels	Utilisation
157'000	18'000	157:18	2 826 000 000	2,826	Better Light Inc., 300 mm lens Digital panorama
3x 6'000 (RGB)					

Distorsions géométriques

Dans une image, les distorsions géométriques se définissent par rapport à la scène. Théoriquement, la correspondance devrait être assurée par un simple changement global d'échelle. Dans la mesure où d'autres changements apparaissent (effet de ballonnement, courbure d'une droite, etc.) on parle de distorsion. Celle-ci peut s'évaluer de façon locale par la distance qui sépare un point idéal du point réel perçu. Plutôt qu'en mm, c'est en pour-cent du champ de vue que cette grandeur est donnée.

Effet de Moiré (recouvrement spectral)

Les capteurs à semi-conducteurs sont formés d'une multitude de cellules qui chacune prélève un "échantillon" de la scène. Mais la technique de l'échantillonnage a ses règles, comme nous l'avons vu dans la partie 1 (§ Information), dont il ressort en particulier que la scène ne doit pas avoir de détails "trop petits". Il est donc souvent judicieux, pour un capteur numérique, de laisser subsister un léger flou lors du réglage de l'objectif. De cette façon l'image tombant sur le capteur ne contient plus de fréquences spatiales trop élevées.

Gain automatique

Les caméras vidéo, à tube ou à semi-conducteurs, disposent en général de circuits automatiques pour le réglage du gain. En présence de signaux lumineux intenses, la sensibilité diminue, et inversement, celle-ci augmente avec l'obscurité.

L'action se passe généralement au niveau de l'électronique de lecture du capteur, mais se prolonge aussi souvent avec un réglage automatique de l'ouverture (diaphragme).

Pour un tout-usage, une variation automatique de gain est très utile. Mais en application industrielle au contraire, on est souvent intéressé à disposer d'un gain fixe, éventuellement paramétrable par l'utilisateur.

La dynamique des systèmes de réglage automatique du gain est impressionnante, et s'étend couramment sur 3 ou 4 décades pour les capteurs à semi-conducteurs.

Temps d'intégration (résolution temporelle)

Le capteur transforme les photons incidents en charges électriques. Durant un certain temps, si les photons continuent d'arriver, la charge totale va augmenter. Ensuite, un équilibre va s'établir, à cause des phénomènes de saturation ou à cause des courants de fuite.

En pratique, il arrive presque toujours que la limite à l'intégration du signal incident soit fixée par le procédé de lecture, qui remet le capteur à l'état "initial" en le déchargeant.

Le temps d'intégration peut aussi être limité par le temps d'ouverture (diaphragme, rideau, élément électro-optique) ou par un signal électrique sur le capteur lui-même (transfert de charges).

Le temps d'intégration varie suivant un compromis toujours différent entre netteté d'un scène en mouvement ou danger de saturation (temps court) et sensibilité de la prise de vue (temps long).

Les performances possibles en termes de temps d'intégration sont nettement différentes entre capteurs linéaires et éléments bidimensionnels. Dans ceux-ci, les temps d'intégration peuvent varier largement. Par exemple 1/1'000'000 s pour une caméra Bosch, à Lonay ou 20 min. pour une application en astronomie, à Genève.

Rémanence

En principe la lecture d'une image devrait être suivie d'une remise "au noir" du capteur, de façon à saisir de façon totalement nouvelle l'image suivante. Il existe des tubes dont les caractéristiques de rémanence sont telles que l'image lue ne disparaît pas complètement. Au contraire, une certaine fraction de l'image perçue à un moment donné continue d'apparaître durant plusieurs lectures ultérieures.

Le phénomène de rémanence n'est en général pas souhaitable. Il peut être parfois utile, comme lors de cas où la scène est fixe, et que le signal perçu est à la fois faible et bruité.

Blooming

Le blooming est un phénomène de diffusion des charges dépêlacées par la lumière incidente.

Ce phénomène peut être visualisé facilement en projetant un point bien net sur un capteur à tube (surface sensible continue). De façon instantanée, il apparaît une charge très "concentrée" sur la surface sensible. Puis ce point s'élargit en une tache de diamètre croissant. On parle parfois d'"éblouissement".

Sur certains capteurs à semi-conducteur, l'éblouissement se produit selon des directions préférentielles. Par exemple un point se propage sur toute une colonne, parce que certains éléments électroniques (amplificateurs, canal de lecture, etc.) y sont communs.

Codage de la couleur

Les capteurs ont des caractéristiques de sensibilité spectrale variable dont on a déjà parlé. Quant au mode de "codage" de la couleur, elle est traitée par ailleurs (§2.9 Composition des couleurs).

En ce qui concerne les capteurs, mentionnons simplement que les caméras couleur actuellement courantes comprennent des éléments sensibles qui sont un par un doté d'un filtre rouge, vert, ou bleu, directement sur la puce lors de la fabrication. De façon similaire à ce qui se passe pour l'oeil humain, il faut donc choisir entre une résolution maximale monochrome, ou une résolution moindre au profit d'un signal coloré.

2.9 Échantillonnage

L'image projetée par l'objectif sur le capteur est continue. Elle ne peut pas être traitée électroniquement telle quelle. Il faut la convertir en un ensemble de petits points adjacents, les échantillons. Ces points sont souvent appelés "pixels" (picture-elements, c'est-à-dire éléments d'images) ou même de façon plus abrégée encore "pels".

La conversion peut s'opérer directement en points (cas des capteurs à semi-conducteur), ou passer par l'étape intermédiaire d'une découpe en lignes (échantillonnage unidimensionnel courant en normes de télévision).

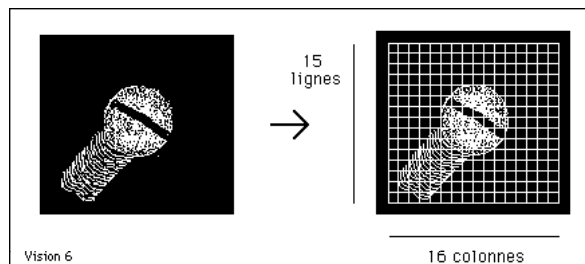


Fig. 2.3 Échantillonnage d'images. L'image continue de gauche est remplacée par 15x16 pixels (une valeur par carré).

Nombre de lignes

Les capteurs à tube ont une surface transductrice continue. L'échantillonnage s'opère selon une dimension (axe vertical) depuis les débuts de la télévision. Jusqu'à récemment, cette norme était très fortement établie. Elle avait la particularité d'être discrète selon une dimension (625, 525, 819 lignes...) tout en restant continue dans l'autre.

En fait cette norme est tellement intéressante par la compatibilité qu'elle garantit entre une immense variété d'appareils, qu'elle a même été adoptée pour la majorité des capteurs à semi-conducteurs. Ainsi ceux-ci, bien qu'ils acquièrent un signal discret selon les lignes, le convertissent en signal continu à leur sortie!

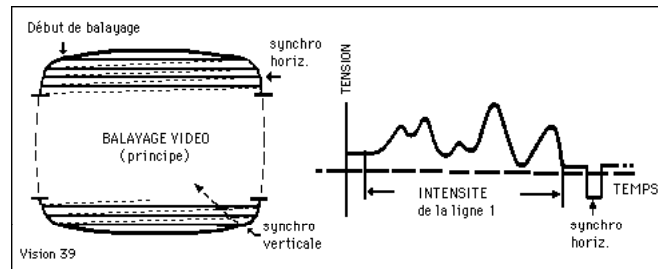


Fig. 2.4 Balayage par lignes d'une image en télévision

La norme la plus courante en Europe assure le balayage de l'image en deux temps. D'abord 312 (et demi) lignes impaires sont parcourues, pour donner une impression rapide de scènes en mouvement, puis les lignes paires, en nombre égal, sont parcourues, de façon à assurer une bonne netteté pour les scènes stables. Il y a 25 balayages complets, soit 50 "trames" par seconde. On parle d'"entrelaçage".

Aujourd'hui, les applications graphiques sur ordinateurs exigent des résolutions plus fines tant dans l'espace (1200 lignes environ) que dans le temps (rafraîchissement complet, c'est-à-dire sans entrelaçage, 50 fois par seconde). Mais ces nouveaux modes sont encore réservés aux moniteurs spécialisés. De façon plus particulière encore, on parle de télévision à "haute définition". Ceci est encore du futur, spécialement si l'on considère l'ensemble de la chaîne allant de l'acquisition d'image jusqu'à sa reproduction, en passant par les équipements de transmission et de stockage.

Nombre de points par ligne

Chaque ligne vidéo reçue doit être convertie en une suite d'échantillons. La cadence indiquée pour le balayage de l'ensemble de l'image implique un rythme très élevé au niveau des échantillons. La cadence est généralement comprise entre environ 8 MHz (256 pixels par ligne) et 16 Mhz (512 pixels par ligne).

Rééchantillonnage

L'échantillonnage a été décrit jusqu'ici de façon quelque peu informelle, en privilégiant l'aspect pratique.

Or cet échantillonnage doit souvent se répéter plusieurs fois, à des résolutions différentes: échantillonnage au niveau du capteur, au niveau du balayage, à celui du numériseur d'entrée de systèmes informatiques. Plus loin encore, le traitement informatique d'images requiert souvent des changements de résolution...

Il faut donc rappeler que la fréquence d'échantillonnage, f_e , doit être légèrement supérieure au double de la fréquence maximale de l'image, f_{max} . Voyons ce qui se passe lorsqu'on s'écarte de cette valeur.

Lorsque f_e est bien supérieure à $2 f_{max}$, il y a gaspillage de temps et de ressources. On traite, transfère et stocke une quantité d'échantillons inutiles. La solution consiste à diminuer f_e .

Lorsque f_e est inférieure à $2 f_{max}$, il convient de décider si les composantes comprises entre $f_e/2$ et f_{max} doivent être retenues. Si oui, alors f_e doit être augmentée. Sinon, il faut éliminer par un filtre passe-bas approprié ces composantes indésirables.

Cette dernière opération est trop souvent négligée, et malheureusement la suite du traitement conduit à des résultats médiocres voire erronés. Il est par exemple courant à la télévision de voir des couleurs irisées sur des habits à motifs fins. Ceci résulte d'une image trop nette pour la résolution imposée par les standards actuels.

2.10 Quantification

Le procédé de quantification consiste à exprimer l'intensité de chaque pixel par un nombre, tiré d'un ensemble fini de valeurs discrètes. La valeur de gris d'un pixel, a priori infiniment variable, ne peut dès lors plus s'écarter d'un certain nombre de "niveaux" de gris.

Voyons d'abord combien de niveaux de gris sont couramment pris en compte. Nous verrons ensuite quels problèmes apparaissent lorsqu'on en retient trop peu, tant en vision par ordinateur que pour l'observateur humain.

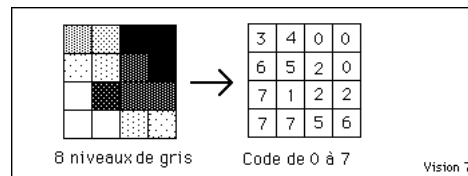


Fig.2.5 Le procédé de quantification remplace l'intensité infiniment variable, de chaque pixel incident, par un nombre.

Nombre de niveaux

Le nombre optimal de niveaux de gris dépend de l'application. Au minimum, deux niveaux sont retenus: blanc et noir. A l'autre extrême, on se heurte au bruit du signal. Pour les caméras courantes, 64 (6 bit) niveaux sont généralement adéquats.

En pratique, les pixels sont souvent décrits par 256 niveaux. Tous ne sont pas significatifs, mais ils sont maintenus car cela peut être décrit par un octet (byte), ce qui est un format extrêmement courant en électronique numérique.

L'oeil humain distingue, semble-t-il, 32 niveaux de gris, répartis logarithmiquement.

Faux contours

Lorsqu'une scène est représentée par un nombre insuffisant de niveaux de gris, de petites variations peuvent disparaître parce qu'elles sont traduites par un même niveau de quantification. Mais le phénomène le plus gênant qui apparaît, c'est maintenant celui des faux contours.

Lorsque la scène présente des surfaces relativement grandes, à faibles variations de gris, ces surfaces apparaissent artificiellement découpées par les frontières entre niveaux discrets successifs.

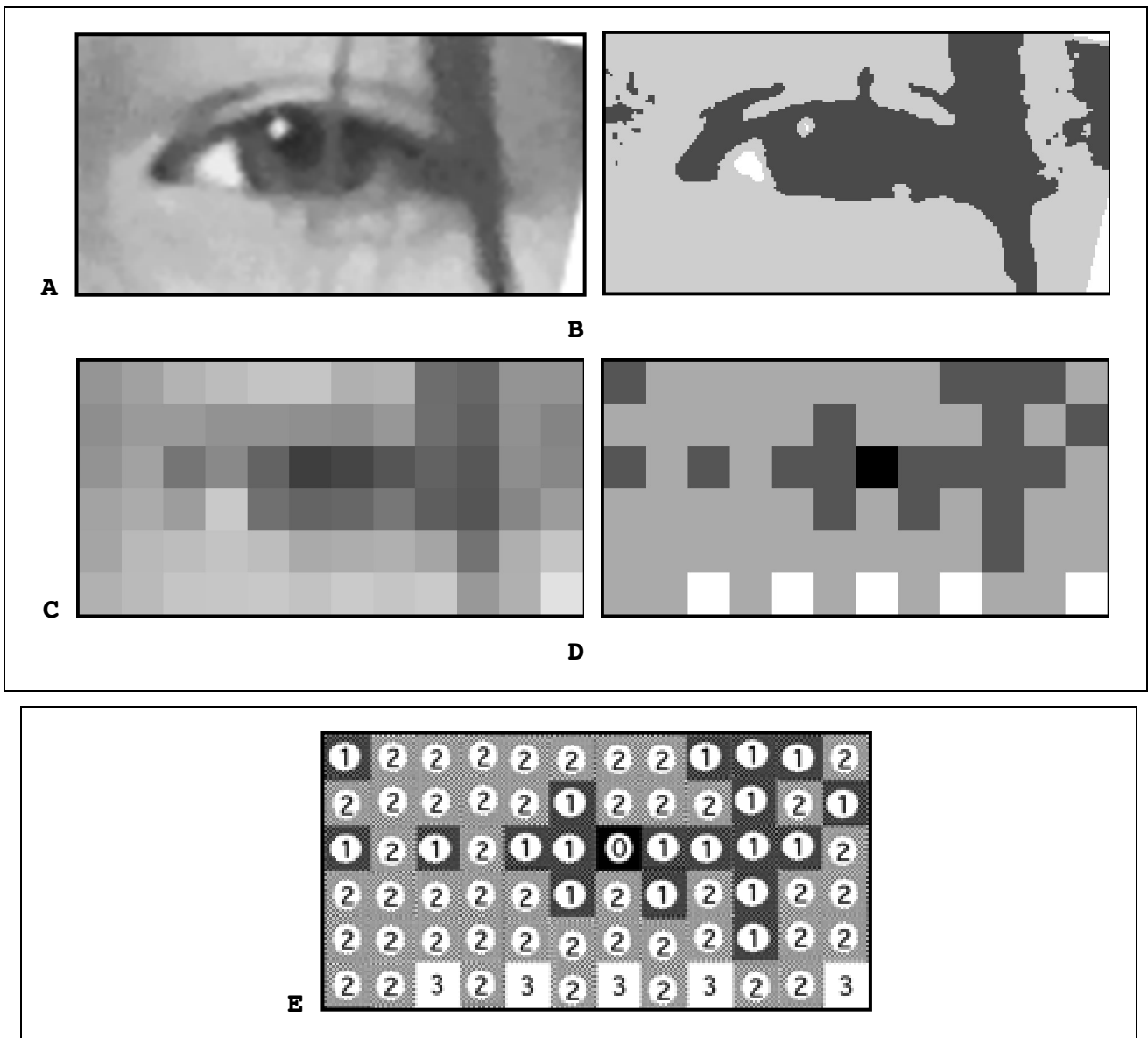
Codage adapté à la vision humaine

L'oeil humain a une particularité que certains types de codage exploitent et qu'il pourrait parfois être souhaitable d'imiter pour une machine: Dans les zones d'une scène à rapides variations d'intensité

(nombreux cycles par mm), l'oeil est peu sensible au niveau. A l'inverse, pour les zones où l'intensité varie peu, les performances en discrimination de niveaux d'intensité sont bien meilleures.

2.11 Exemple de numérisation d'image

Dans ce paragraphe, un exemple complet de numérisation est présenté. Sur la fig. ci-dessous, l'image originale (très légèrement floue pour respecter le théorème d'échantillonnage) est échantillonnée de façon grossière puis quantifiée à 4 niveaux de gris et enfin codée par les valeurs 0 à 3. A titre indicatif, la fig. présente aussi le résultat d'une quantification directe sur la première image, mais ceci n'est en général pas requis.¹



¹ En fait ici tout est numérisé mais ce qui est fait à grande résolution (300 pixels par pouce) peut être assimilé à du continu.

Fig. 2.6 En haut: l'"original" (A), et sa quantification à 4 niveaux (B)
 Au milieu: échantillonnage grossier (C), et sa quantification à 4 niveaux (D)
 En bas (E): code de 0 à 3 décrivant les pixels de l'image D, superposé sur celle-ci.

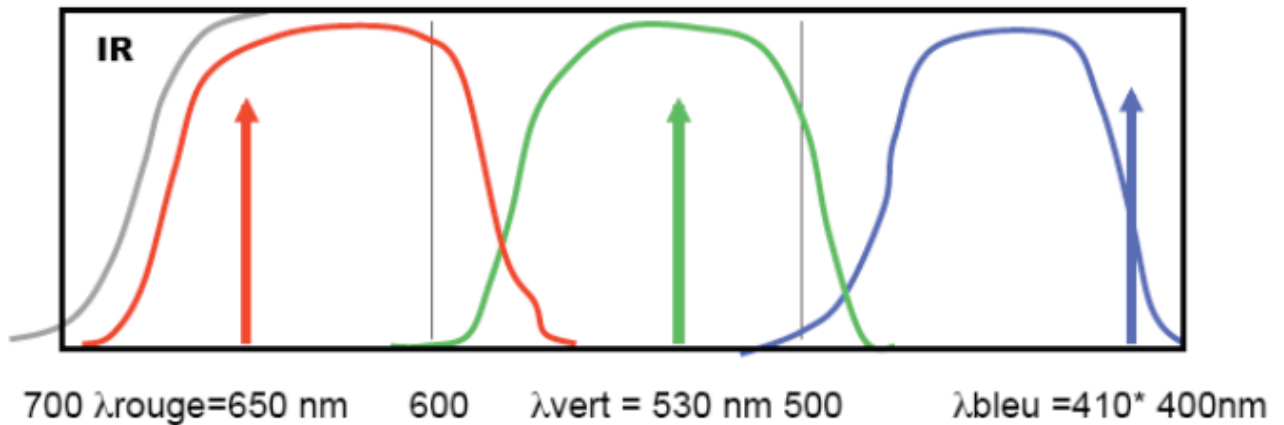
2.12 Composition des couleurs

La plupart du temps, c'est l'intensité lumineuse d'un pixel qui est intéressante. Mais parfois, il est intéressant de prendre en compte ses caractéristiques spectrales. Il peut s'agir d'infrarouge, ou d'ultraviolet. Si l'intensité est dans le visible, elle peut encore être d'une couleur particulière.

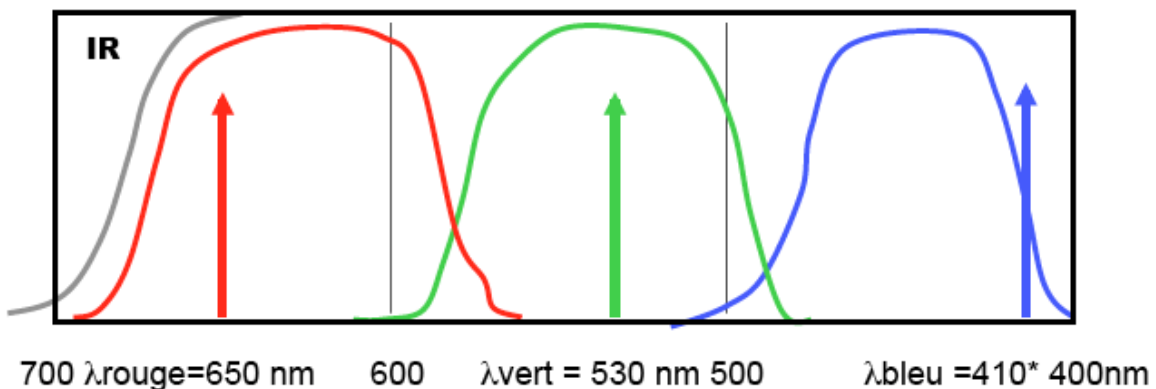
L'expérience montre que la variété presque infinie de couleurs que l'homme perçoit peut en fait se représenter par une combinaison de trois couleurs de base. Les couleurs habituellement retenues ont les longueurs d'onde suivantes:

$$\lambda_{\text{rouge}} = 650 \text{ nm}, \lambda_{\text{vert}} = 530 \text{ nm}, \lambda_{\text{bleu}} = 410 \text{ nm}$$

En télévision, lorsqu'on perçoit une scène, on en distribue le contenu en trois bandes spectrales adjacentes, centrées sur les 3 longueurs d'ondes de base, à l'aide de filtres appropriés.



* HP-Microsoft (sRGB): R:633, V:514, B:442 nm



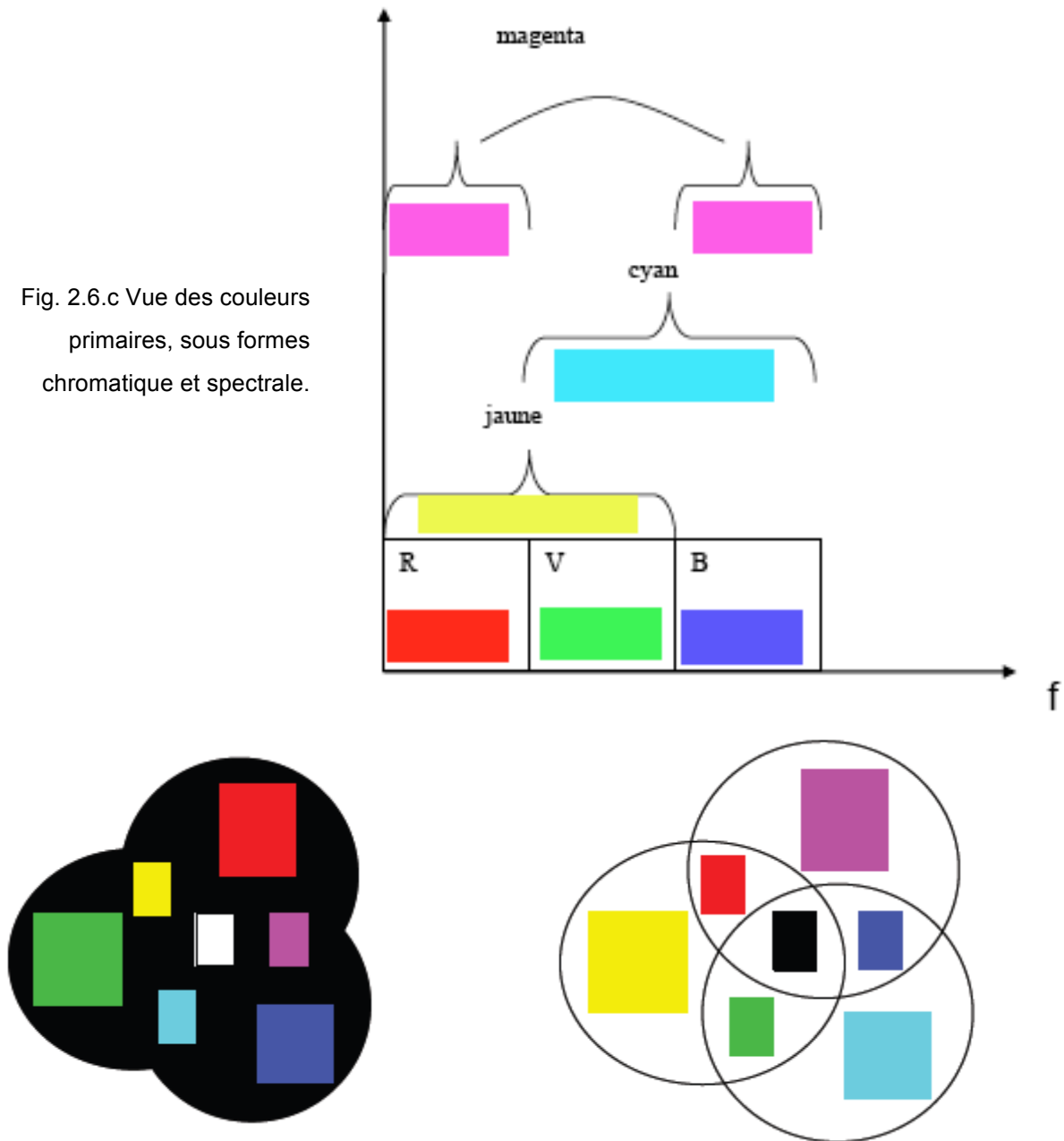
* HP-Microsoft (sRGB): R:633, V:514, B:442 nm

Fig.

2.6.b Bandes passantes (principe) et longueurs d'ondes standard pour la gestion tridimensionnelle des

couleurs en mode RVB. Noter le filtre bloquant les infrarouges (gris, à gauche) requis pour obtenir de belles couleurs à l'aide d'un capteur à semiconducteur.

Fig. 2.6.c Vue des couleurs primaires, sous formes chromatique et spectrale.



. 2.6.c Synthèse additive ($R+V+B=\text{blanc}$), à gauche, et soustractive ($\text{blanc}+(\text{jaune}=-B)+(\text{cyan}=-R)+(\text{magenta}=-V)=\text{noir}$), à droite

En analyse d'images, on peut hésiter entre l'utilisation générique de ce type de représentation des couleurs, et l'utilisation de filtres spécifiques, tirant parti des aspects uniques d'une application donnée. Dans ce dernier cas, on pourra par exemple distinguer des longueurs d'ondes qui sont habituellement confondues, traversant le même filtre rouge (R), vert (V, ou en anglais, G), ou bleu (B).

Intuitivement, les couleurs ne varient pas en fonction de l'intensité. De même, on peut introduire des signaux normalisés par rapport à l'intensité n_r , n_v , et n_b :

$$n_r = \frac{R}{R+V+B}; \quad n_v = \frac{V}{R+V+B}; \quad n_b = \frac{B}{R+V+B};$$

Notons que dans ce cas la somme des trois signaux est toujours égale à 1. En conséquence, nous pouvons décrire la couleur normalisée avec seulement deux variables. Ceci est exploité pour créer des planches (2D) de bien des couleurs possibles (Fig. 2.7).

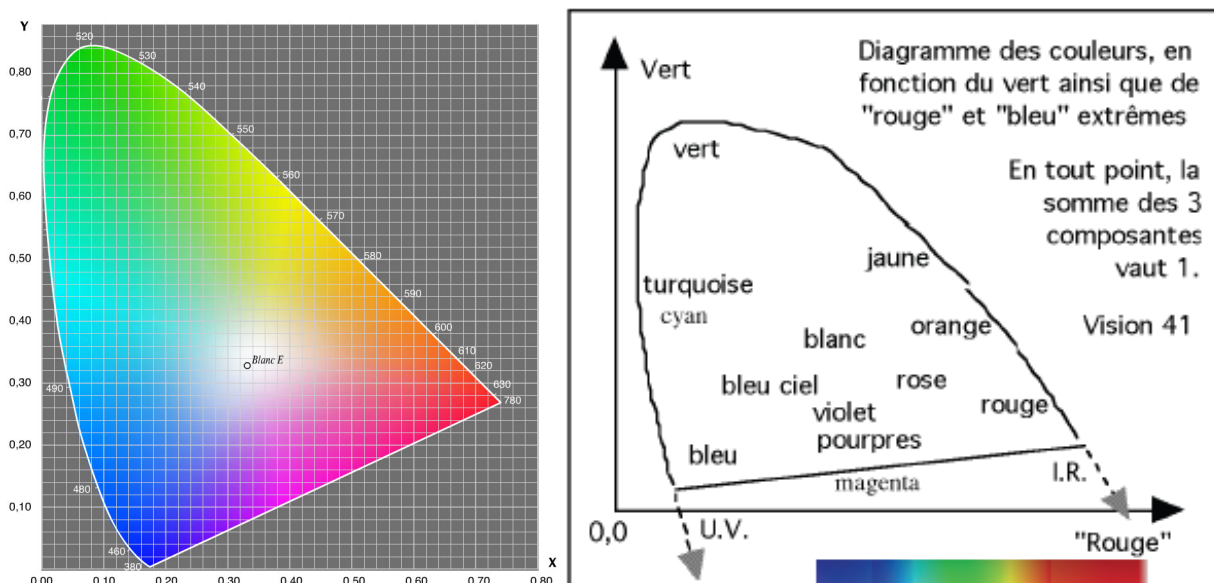


Fig. 2.7 Diagramme "en langue" des couleurs, selon norme de la Commission Internationale de l'Éclairage CIE.

Remarques relatives au diagramme "en langue" du standard couleur défini pour la télévision (C.I.E. Chromaticity Diagramm).

1. Coordonnées ; en abscisse : intensité à la sortie du filtre rouge normalisé et en ordonnée, sortie du filtre vert normalisé
2. La teinte est une couleur « pure » correspondant au bord du diagramme CIE. Sur deux côtés on a le spectre de la lumière (signaux monochromes de longueur d'onde donnée) et sur le troisième côté (bas), la droite des pourpres.
3. La saturation caractérise la part de blanc dans une couleur. Si la saturation est 0, on est sur le point blanc. Si au contraire c'esr 1, on a une teinte pure.
4. La figure montre seulement les couleurs dont la norme d'intensité vaut « 1 ». Par exemple le gris (intensité : 0.5) ou le brun (c'est du jaune foncé : intensité = 0.5 dans le rouge et dans le vert, 0 dans le bleu)
5. En modulant trois sources ponctuelles, on peut se déplacer à l'intérieur de ces espace seulement, soit un triangel dans le diagramme de la CIE

6. La lumière émise par un corps noir se caractérise par un point qui se trouve sur une courbe arrondie expérimentale allant du rouge pur (température de quelques centaines de degrés) jusque dans le bleuté (>5000 degrés) en passant par le blanc. Par contre intuitivement on par le couleurs « chaudes » vers le rouge et de couleurs »froides » dans les bleus.
7. En traitement d'images, le vocabulaire spécialisé fait correspondre le bleu ciel, à un bleu moyennement saturé. Le rose à un rouge moyennement saturé, le cyan à l'addition de vert et de bleu, et le magenta à un mélange additif de rouge et de bleu. Dans le langage courant on confond souvent rose et magenta, bleu ciel et cyan, et enfin pourpre, violet, ou magenta ne se distinguent pas.

Notons que le diagramme de la CIE est basé sur les caractéristiques de l'homme (voir fig. 2.7b).

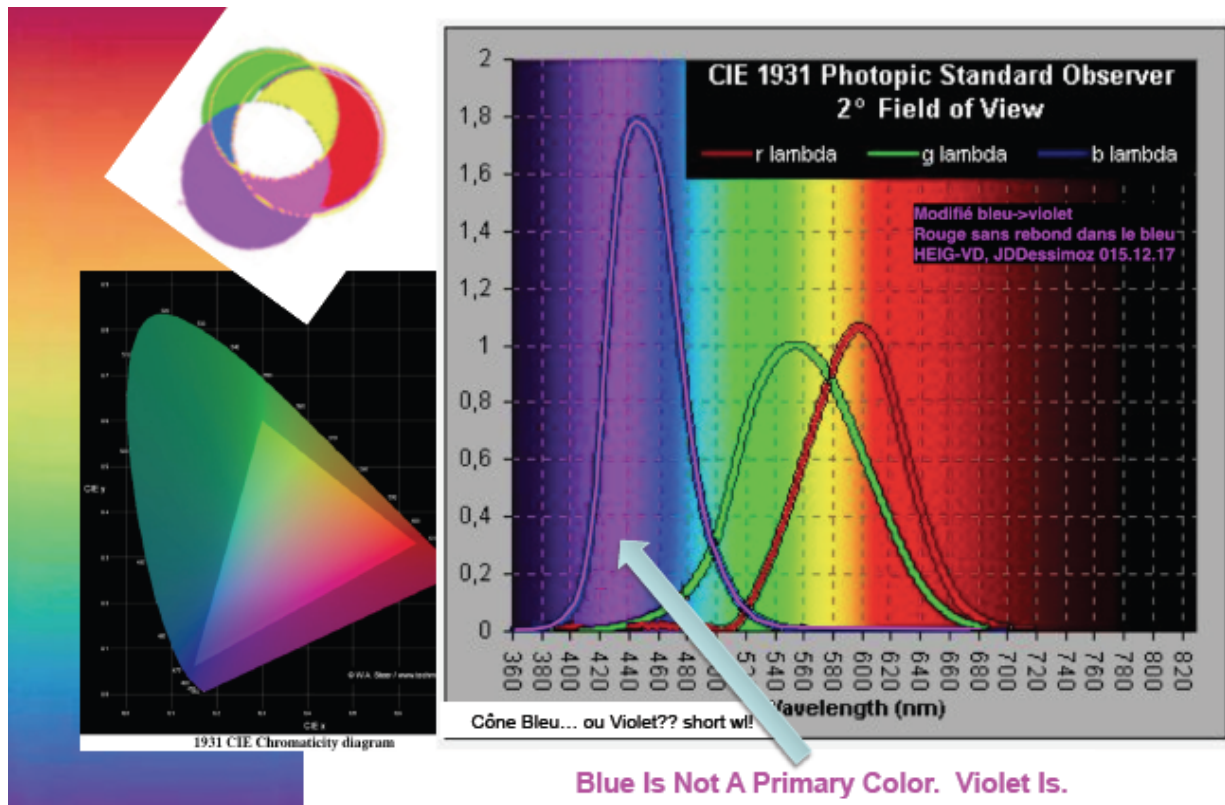


Fig. 2.7b Caractéristiques de sensibilité de l'œil humain, à l'origine du diagramme classique de la CIE, presque centenaire et présenté en fig.2.7 (doc. www.aim-dtp.net 2006). Alternative : les cônes sensibles aux longueurs d'ondes courtes sont en fait sensibles au violet plutôt qu'au bleu (dès lors le rebond dans le rouge peut s'ignorer), <http://www.sjsu.edu/faculty/watkins/violet.htm> !

Il est parfois utile de décomposer le signal de couleur non pas en composantes R, V, B, mais selon trois autres critères qui mette en évidence d'autres grandeurs physiques: L'intensité lumineuse, I, la couleur "intrinsèque" ou la teinte, T (ou H comme "hue" en anglais), et le niveau de saturation S.

La première caractéristique, I, est la plus simple. C'est l'intensité lumineuse globale, que l'on peut retrouver dans le cas des signaux RVB par leur addition.

La deuxième composante, c'est la couleur "intrinsèque", ou teinte. Avant d'aller plus avant, rappelons que le noir c'est l'absence de tout signal (intensité nulle), le blanc, c'est l'intensité maximale sur l'ensemble de tout le spectre. Et le gris représente une situation intermédiaire de l'intensité, mais toujours pour un spectre constant (en particulier, même intensité perçue dans chaque canal RVB). La composante qui nous intéresse ici, c'est précisément celle qui rend compte du déséquilibre entre bandes spectrales particulières, qui se manifeste subjectivement par la perception de couleurs.

Enfin la grandeur de saturation indique combien la couleur intrinsèque est pure ou au contraire "diluée de blanc". En peinture on va, par addition de blanc, faire évoluer une couleur marquée (par exemple rouge) vers un ton pastel (rose). A partir des composantes RVB, les composantes I et S peuvent se déduire comme suit:

$$I = R + V + B$$

$$S = \frac{I - \text{Min}(R, V, B)}{I}$$

Quant à la composante de teinte, T, elle traduit la position de la teinte dominante sur le spectre lumineux (entre le rouge et le bleu), ou sur la "droite des pourpres" (cf. fig. 2.7). La droite des pourpres correspond aux couleurs ne contenant aucun vert. On la parcourt en remplaçant progressivement le bleu par le rouge.

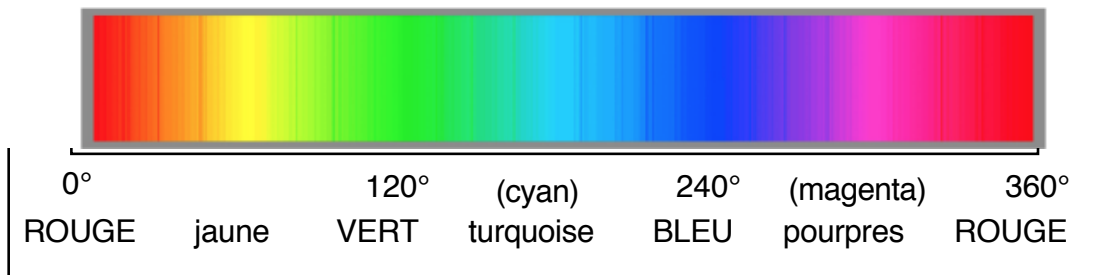


Fig. 2.8 Codage de la teinte par un angle

Du point de vue métrique, on peut rendre compte de la teinte T par la méthode de principe suivante (cf. aussi fig. 2.8):

1. Considérons, parmi les trois composantes de couleurs R, V et B les deux plus grandes.
2. Ceci nous donne de façon directe l'intervalle R-V, V-B, ou B-R où T se trouve
3. Puisque les point de départ et d'arrivée sont confondus (R), il est logique d'associer la teinte, T, à une variable angulaire allant de 0 à 2π.
4. Dès lors le domaine R-V couvre l'intervalle {0..2π/3}, V-B devient {2π/3..2π*2/3}, et B-R, {2π*2/3..2π*3/3}.
5. Et la position fine de T à l'intérieur de l'intervalle se trouve en prenant en compte le poids relatif des deux composantes retenues.

Ainsi par exemple si $R=0.1$, $V=0.2$, et $B=0.4$, alors $I=0.1+0.2+0.4=0.7$; $S= (0.7-0.1)/0.7$; les deux plus grandes composantes sont V et B , valant respectivement, en fonction de la troisième composante, $V'=0.2-0.1=0.1$ et $B'=0.4-0.1=0.3$. Dès lors,

$$\begin{aligned}
 T &= V'/(V'+B') \cdot 2\pi/3 + B'/(V'+B') \cdot 2\pi \cdot 2/3 = \\
 &= 0.1/(0.1+0.3) \cdot 2\pi/3 + 0.3/(0.1+0.3) \cdot 2\pi \cdot 2/3 = \\
 &= 1.75 \cdot 2\pi/3
 \end{aligned}$$

Il existe encore d'autres systèmes de compositions que les RVB et ITS présentés ci-dessus. En particulier, nous mentionnerons encore les systèmes YUV et C-M-Y-K. Le mode YUV est intéressant pour la compatibilité avec le noir/blanc. Y est la somme des trois canaux RVB. Quant aux deux autres composantes, elles forment le signal de chrominance: $U=Y$ -Rouge, et $V=Y$ -Bleu. Enfin, dans l'industrie graphique, il apparaît pratique d'utiliser pour les noirs, de l'encre noire, plutôt que la combinaison de trois encres colorées. On a donc les composantes de couleurs C (cyan), M (magenta), Y (Y n'a pas ici la même signification que plus haut; ici il est l'initiale de "yellow", c'est-à-dire jaune). Enfin K désigne le noir.



Visible light



Near infrared "Night Vision"



Thermal infrared

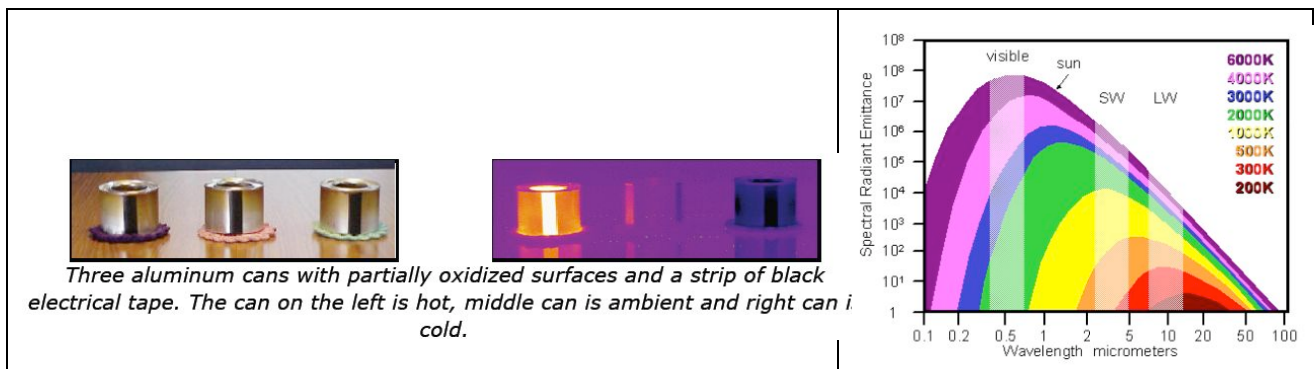


Fig. 2.9 Images (cf. ref. bibliographique « Infrared ») dans l'infrarouge, avec caméras similaires à ce qui se fait dans le visible, et avec microbolomètres. Il faut observer que le signal reçu dépend non seulement de la température mais aussi de l'émissivité des surfaces. De façon plus générale, l'image peut résulter d'autres phénomènes physiques encore, tels que rayons gamma ou balayage de barres à ultrasons

2.13 Marche à suivre pour acquérir une image

Lorsqu'on saisit une image pour une application en vision par ordinateur, il faut prendre un grand soin à respecter les lignes directrices rappelées ci-dessous:

- utiliser les grandeurs *physiques* les plus discriminantes pour faire ressortir la partie intéressante de la scène (le "signal") par rapport à tout le reste, qui est plutôt gênant dans l'application donnée (le "bruit"); les grandeurs physiques peuvent être très différentes de la lumière : vitesse, hauteur, épaisseur, magnétisme, température, nature du matériau, etc.
- passer des grandeurs physiques discriminantes à des phénomènes *lumineux*; par exemple, une différence de hauteur peut être mise en évidence par un éclairage latéral, et donc par la présence d'ombres.;
- et limiter la quantité d'information saisie au minimum requis pour l'application

Augmentation de contraste

Pour augmenter le contraste entre éléments d'intérêt et éléments gênants, on peut, durant l'acquisition d'image, recourir à de nombreuses caractéristiques physiques. Souvent ce sera simplement la quantité de lumière réfléchie, qui fournira un indice suffisant, mais très souvent également, on gagnera à considérer d'autres "dimensions": couleur, transparence, fluorescence, température, vitesse, etc. (voir tableau 2.1).

Par exemple, on a eu mis en évidence des niveaux d'eau, dans des bouteilles semi-transparentes, en "jouant" sur l'indice de réfraction différent entre air et eau. Ou dans d'autres applications, on a recouru à la fluorescence de certains tissus vivants, dans une pomme de terre, pour localiser facilement la base des germes...

Grandeurs <i>physiques</i> discriminantes	Actions envisageables à la prise de vue
Localisation	cadrage
Distance à l'objectif	focale, profondeur de champ
Grandeur géométrique	Filtrage passe-bas spatial (flou)
Anisotropie directionnelle	Lentille cylindrique (focale différente selon deux axes perpendiculaires)
Orientation des surfaces	Angles d'éclairage ou de prise de vue
Vitesse	Temps de prise de vue (éclairage stroboscopique, sensibilisation du capteur, etc.) Suivi de l'élément mobile Moyennage d'images Soustraction d'éléments fixes
Couleur, longueur d'onde	Filtre couleur, source de lumière et capteur appropriés

Indice de réfraction	Mise au point, position de la source lumineuse ou du capteur
Réflexions	Filtre polarisants, direction d'éclairage et de prise de vue
Fluorescence	Eclairage et capteur adaptés aux longueurs d'ondes en jeu
Réflectance	Eclairage épiscopique
Caractéristique d'opacité	Eclairage diascopique

Tableau 2.1 Utilisation de différences physiques pour obtenir un bon contraste

Grandeur physique	Action envisageable
<i>Espace (géométrie)</i> Etendue du champ de vue (taille de l'image)	- Cadrage serré!
Résolution spatiale	- Netteté minimale! - filtrage passe-bas (flou) et échantillonnage basse fréquence ($F_{éch} \approx 2F_{max}$) - Analyse séparée selon 2 axes perpendiculaires
<i>Temps</i> Durée (nombre d'images)	En général: instantané (prise de vue unique)!
Résolution temporelle	- $F_{éch} \approx 2F_{max}$! - Filtrage passe-bas (temps d'ouverture approprié) éventuellement nécessaire.
<i>Intensité lumineuse</i> Gamme dynamique	- Exploiter toute la plage de sensibilité du capteur: du minimum (niveau de bruit; noir) au maximum (niveau de saturation; blanc)
Résolution des niveaux de gris	- Nombre minimal de nuances. - Viser 2 niveaux (noir-blanc)
<i>Spectre lumineux</i> Résolution des couleurs	- Nombre minimal de couleurs / longueurs d'ondes différentes - Essayer d'éviter les couleurs (travailler sur l'intensité; luminance). Cf. niveaux de "gris".

Tableau 2.2 Limitation de la quantité d'information au minimum significatif

Réduction de la quantité d'information à saisir

Il ne suffit pas d'avoir un bon contraste dans la scène. Il faut aussi faire en sorte de limiter la quantité d'information à traiter à la plus petite valeur possible. Pour cela on a la double démarche, pour chaque

paramètre significatif, de limiter la dynamique (l'étendue) au minimum, et de garder une résolution aussi grossière que possible.

Par exemple, on ne retiendra d'une image, pour une application donnée, que 20 lignes, espacées de plusieurs millimètres. Ou encore on ne prendra en compte que 5 couleurs possibles. Etc. (voir Tableau 2.2).

Quelques exemples suivent.

<p>Points balayés</p>	<p>(source : document allemand)</p>
<p>Opacité donc diascopie (Conti et Zanni 2008)</p> <p>Hauteur donc éclairage oblique (Conti et Zanni 2008)</p> <p>0: noir, 1-30: bleu, ... cyan... 128: vert, ... jaune, 220-254: rouge, 255: blanc Pseudocouleurs</p>	<p>Couleur (filtre)</p> <p>Ecran fausses couleurs</p> <p>Ecran niveaux de gris</p> <p>Palet rouge</p> <p>Palet vert</p> <p>Phosphorescence (Eclairage UV)</p> <p>Vitesse, déplacement (stroboscope, demi-balayage)</p>
<p>sachets vide et plein*</p>	<p>Acquisition en transparence pour sucre (gauche) et vue de Kern en modes visible et infrarouge (ci-dessus)*</p>

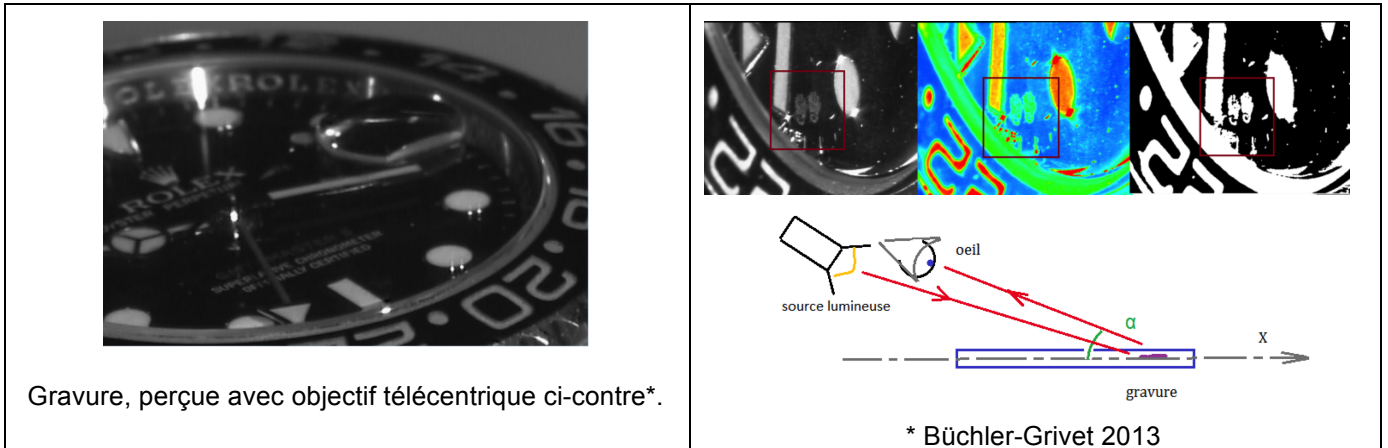


Fig. 2.13 (début) Quelques exemples d'images « brutes » ou « bien acquises »

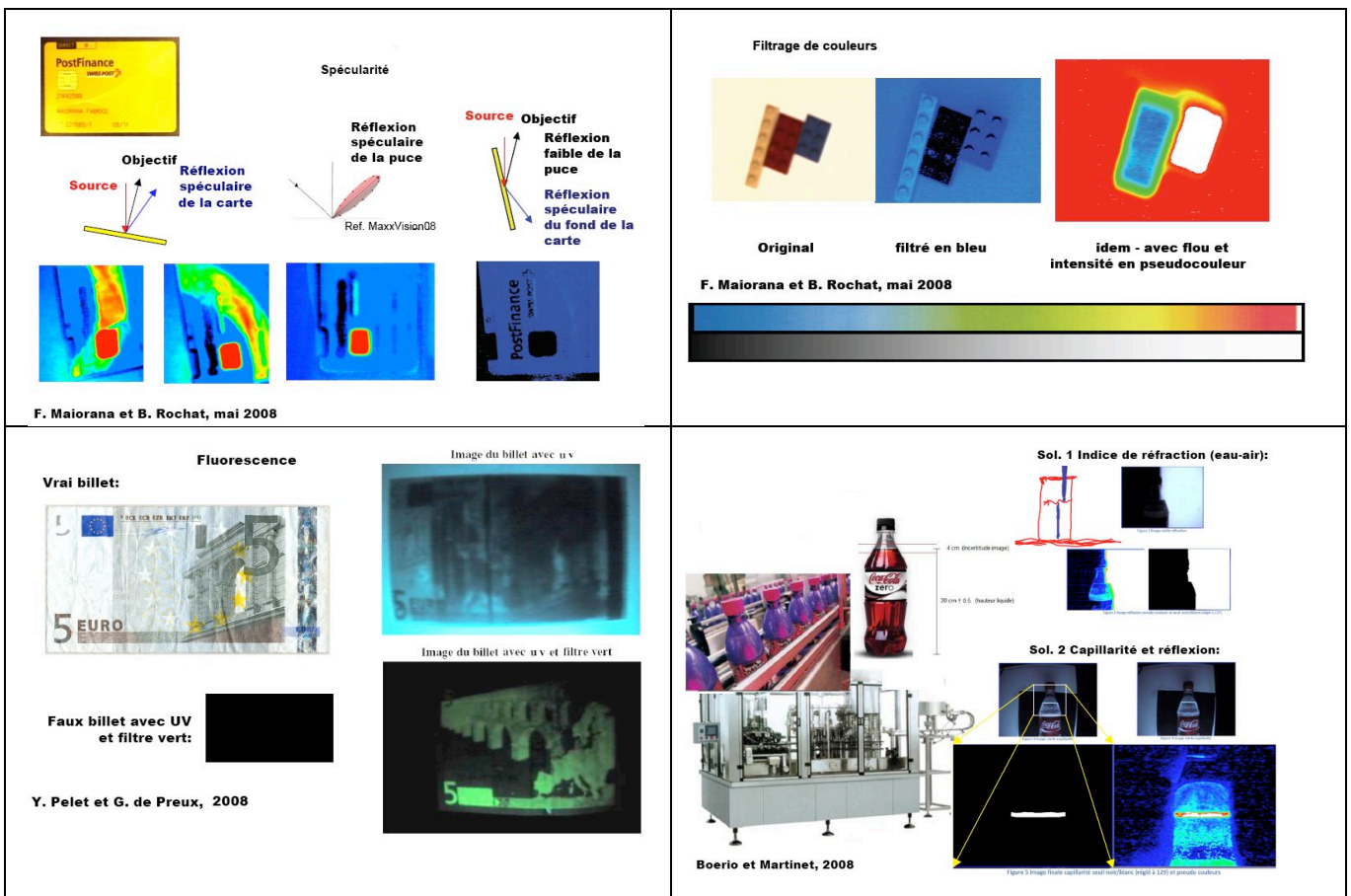


Fig. 2.13 (suite) Quelques exemples d'images « brutes » ou « bien acquises »

3 TRAITEMENT BIDIMENSIONNEL D'IMAGES

Le traitement d'images peut se comprendre de façon très générale et englober l'ensemble de la vision artificielle. Ce n'est pas dans ce sens que nous l'utilisons ici. Au contraire, il a un sens très limité, puisqu'il ne recouvre que les opérations ayant aussi bien en entrée qu'en sortie une structure

bidimensionnelle. Exemple: avec cette acception restrictive, l'inversion d'une image (calcul du négatif) appartient au traitement bidimensionnel d'images; mais l'estimation du gris moyen de l'image ne le concerne pas.

Examinons le traitement dans des voisinages de taille croissante. D'abord, le traitement se fait de façon identique en chaque pixel considéré isolément. Ensuite, quelques points voisins seront pris en compte dans la mise à jour de chaque pixel de sortie. Enfin, c'est le travail à l'intérieur de toute une région ou fenêtre, voire même sur l'ensemble de l'image qui sera discuté.

3.1 Opérations par pixels individuels (point par point)

Les opérations concernant chaque pixel considéré individuellement sont les plus simples à définir et à exécuter. Voyons d'abord le principe de ces opérations, puis leurs applications particulières en quantification, en "amélioration" d'images et pour l'affichage en pseudo-couleurs. Remarquons que ces deux derniers traitements n'ont pas de sens en analyse d'images par machines, mais s'adressent purement à l'observateur humain.

Fonctions de transfert

Les opérations sur pixels individuels peuvent être variées mais ont en commun une très grande simplicité d'utilisation puisqu'elles peuvent se faire sans aucun calcul. Un pixel ne pouvant avoir qu'un nombre limité d'états (typiquement 256 au maximum), il est facile de tabuler la fonction qu'on veut leur appliquer.

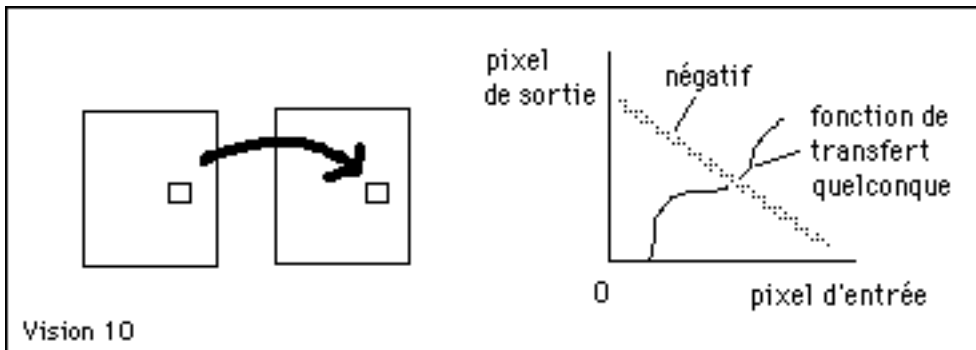


Fig. 3.1 Les opérations sur pixels peuvent se tabuler facilement.

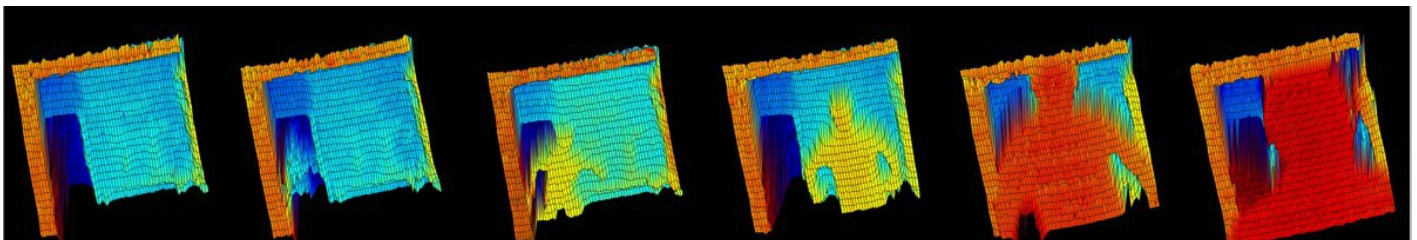


Fig. 3.1bis La distance est représentée par des pseudo-couleurs par un traitement pixel par pixel (photo CSEM).



Fig. 3.1ter Modification de la fonction de transfert des pixels sur Photoshop.

Le remplacement de la valeur d'un pixel par le correspondant en mémoire peut se faire à des cadences extrêmement grandes, qui en pratique ne sont jamais limitées. Ce qui prend plus de temps, c'est le chargement de la table de conversion ("Look-up table"). Mais même ceci est rapide, et peut même être masqué par l'usage de plusieurs tables de conversion.

Une idée similaire est celle des palettes de couleurs, où une table contient de façon détaillée (typiquement, sur 24 bit) la description des couleurs retenues dans la palette (typiquement 256 ou 32000, mais parfois beaucoup moins: 4, 8...).

Seuillage, quantification

Une table de conversion peut s'utiliser pour changer le nombre de niveaux de quantification. Il est ainsi possible de réduire le nombre de niveaux, mais bien sûr pas de l'augmenter.

Si l'on veut par exemple passer de 256 niveaux à 2, on chargera par exemple la valeur 0 dans la table de conversion pour les adresses 0 à 127, puis 1 (ou 200...) pour les valeurs allant de 128 à 255.

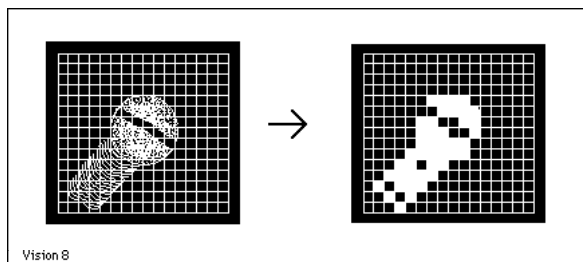


Fig. 3.2 La figure tente d'illustrer le passage de n niveaux, où n est grand, à 2 niveaux.

Amélioration d'image

Il arrive qu'une image soit subjectivement mauvaise: trop claire, trop foncée, trop contrastée, etc. Ces phénomènes peuvent être automatiquement mis en évidence par analyse de l'histogramme des niveaux de gris (cf. §4.1.A).

Tous ces cas peuvent être traités par table de conversion. Prenons l'exemple d'une image trop claire. L'histogramme indiquerait que les pixels ont leurs valeurs concentrées entre 150 et 250. On pourrait alors charger la table de conversion avec 0 pour les adresses allant de 0 à 150, puis par $5 \cdot (\text{pixel} - 150)$ dès cette valeur. Une telle transformation "étirerait" les niveaux de gris entre le noir et le blanc, ce qui devrait être subjectivement plus agréable. Il est clair que du point de vue de la quantité d'information contenue dans l'image, rien ne s'est amélioré, et qu'un tel traitement n'a de sens que si le destinataire de l'image "améliorée", c'est l'oeil humain.

Pseudo-couleurs

L'oeil est également très sensible aux couleurs. Il est courant, tant pour l'impact émotionnel que pour coder certaines caractéristiques physiques que l'on recoure à un affichage en couleur.

Le codage repose alors presque toujours sur l'utilisation de tables de conversion séparées pour les sorties rouge, vert, et bleu d'un numériseur d'image. Si l'on désire par exemple afficher en rouge l'état "15" d'un pixel, on chargera une valeur maximale à l'adresse 15 de la table du canal rouge (R), et la valeur minimale, 0, à l'adresse 15 des deux autres tables de conversion, V et B.

3.2 Modifications des pixels d'après leur voisinage

Alors que tous les numériseurs sont capables de réaliser matériellement les opérations point-à-point, le traitement d'après un voisinage est beaucoup plus incertain.

Il s'agit ici principalement de filtrer spatialement des images, de façon linéaire ou non. Mais il existe aussi certaines méthodes dites de relaxation que nous verrons plus bas, ainsi que des traitements à résolution variable ("scale-space").

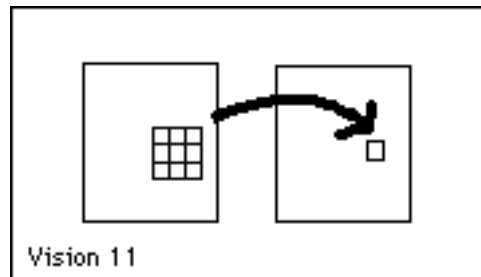


Fig. 3.3 Principe du traitement bidimensionnel tenant compte du voisinage de chaque pixel.

Ce paragraphe se termine par une brève discussion de la texture.

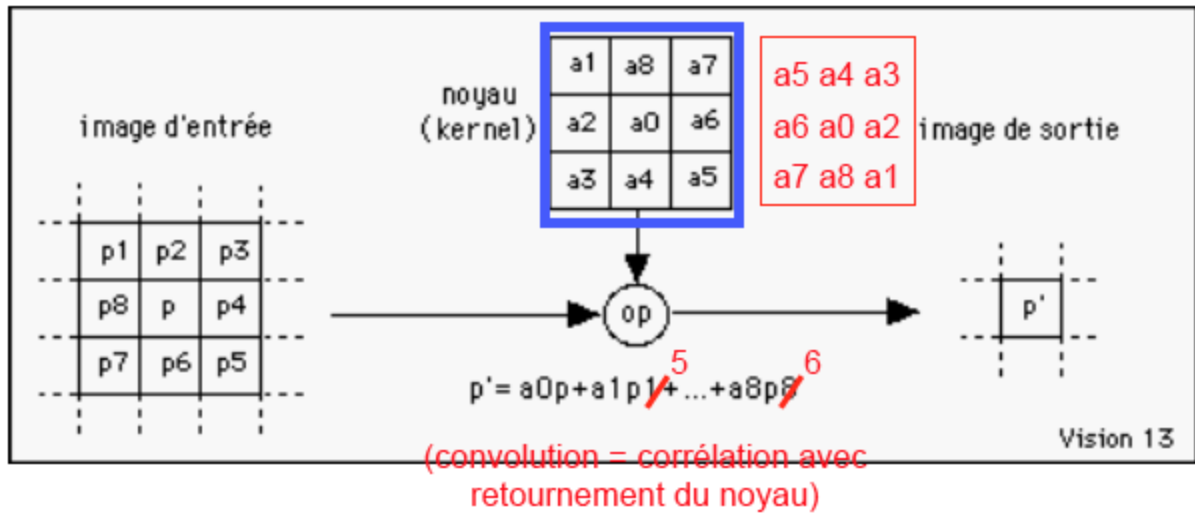


Fig. 3.4 Principe du filtrage linéaire d'une image. Le noyau est appliqué, successivement ou en parallèle, à tous les points de l'image

3.2.1 Filtres linéaires

En traitement du signal, les filtres se décrivent par leur réponse à une impulsion. Dans le cas des images, les filtres travaillent dans le domaine des fréquences spatiales, et leur réponse impulsionnelle se décrit par un "noyau" (en anglais "kernel"). L'"impulsion" dont on parle est ici une image noire contenant un pixel blanc en son centre.

La fig. 3.4 décrit le principe de l'opération de filtrage, que l'on peut aussi exprimer par une équation:

$$I'(l, c) = \sum_{i=n/2}^{n/2} \sum_{j=m/2}^{m/2} k(i, j) * I(l - i, c - j)$$

Cette expression décrit l'opération dite de "convolution" que le filtrage réalise. Il est possible aussi parfois de substituer, selon le même principe, une série de filtres "simples" à un filtre "compliqué".

$$\begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} * \begin{bmatrix} -1 & 0 & 1 \end{bmatrix} ; \quad \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} * \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

Fig. 3.5 Un filtre linéaire "compliqué" peut parfois se remplacer par une cascade de filtres plus simples. L'"*" indique l'opération de filtrage (convolution).

Les filtres que nous allons discuter seront successivement un filtre passe-bas, un passe-haut, et un filtre adapté.

Moyenneur

L'exemple type de filtre passe-bas est le moyenneur. A sa réponse carrée (en amplitude) dans le domaine spatial, correspond un sinus x/x dans celui des fréquences. La bande passante est minimale pour une longueur donnée du filtre. Par contre les lobes latéraux sont assez marqués.

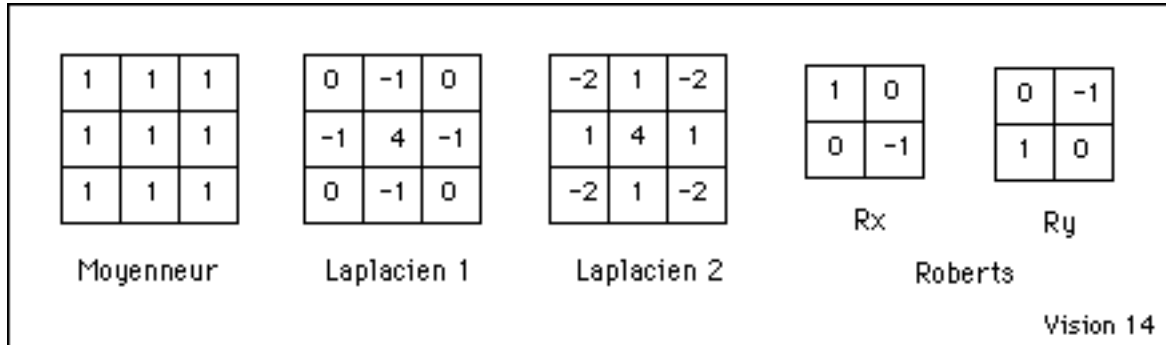


Fig. 3.6 Filtres linéaires passe-bas (moyenneur) et passe-haut (Laplaciens). Souvent, les filtres sont "décomposables" et traiter indépendamment 2 directions indépendantes (ex.: filtre de Roberts)

En général, ce filtre est applicable de façon à ce que la quantité de calcul ne dépende pas de la grandeur de la réponse impulsionnelle, ce qui contribue à le rendre attractif.

Souvent, une composante de type moyenneur est présente dans un filtre différent par ailleurs.

En général, une grande réponse impulsionnelle ne se justifie guère car cela implique une forte perte de contenu spectral qui est contradictoire avec le choix fait de la résolution du capteur.

Notons qu'une simple défocalisation de l'objectif provoque un filtrage moyenneur de façon optique.

Extraction de bords

Les bords se caractérisent, dans une image, par un changement rapide de l'intensité lumineuse. Ces changements peuvent être mis en évidence par une opération de dérivation, ou, de façon plus générale, par filtrage passe-haut.

L'information semble se concentrer, à l'intérieur d'une image, sur les zones de changement de gris. Tant les expériences sur la perception de l'homme, en psychologie, que les équations de base, en théorie de l'information, mettent en évidence ce phénomène.

La fig. 3.6 présente plusieurs filtres passe-haut. Le premier d'entre eux (Laplacien 1) implément la double dérivée de l'intensité lumineuse par rapport aux déplacement spatiaux. Le deuxième s'en distingue par "l'addition" d'une composante passe-bas (mathématiquement, si deux filtres sont appliqués successivement, la réponse impulsionnelle d'un filtre unique équivalent est la "convolution" des réponses impulsionnelles des deux filtres). A l'inverse, les filtres sont parfois "décomposables" et un filtre unique peut être remplacé par deux filtres plus simples qui traitent indépendamment 2 directions indépendantes (ex.: filtre de Roberts).

Les deux noyaux de Roberts permettent de calculer à la fois l'intensité, I , et la direction, θ , des éléments de bords d'une images:

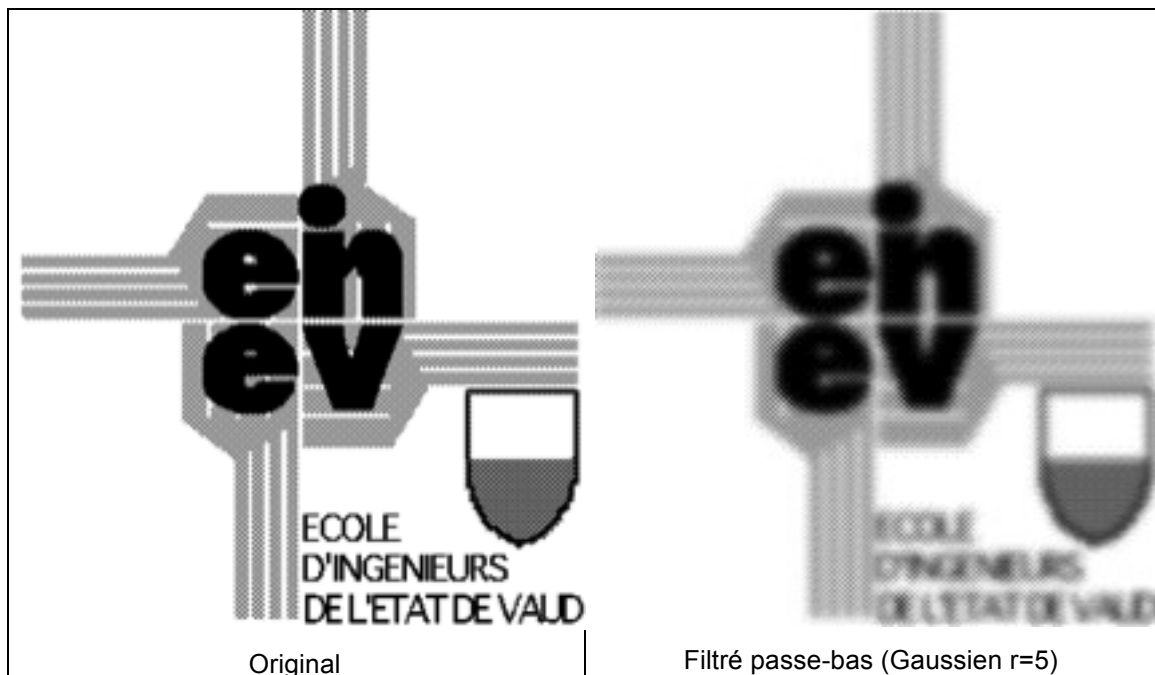
$$I = \sqrt{R_x^2 + R_y^2} \quad ; \quad \theta = \text{Arc tg} \left(\frac{R_y}{R_x} \right)$$

La littérature scientifique propose un nombre incroyablement élevé de filtres extracteurs de bords. La différence entre eux est probablement négligeable car de toute façon leur réponse impulsionnelle doit être échantillonnée avec un nombre extrêmement réduit d'échantillons (typiquement 3, pour une direction donnée!).

En général, la démarche devrait être la suivante: évaluation des détails significatifs de la scène, pour une application donnée. De là on tire la fréquence de coupure du filtre passe-bas (optique) nécessaire, ainsi que la résolution d'échantillonnage. A ce stade, si des bords doivent être mis en évidence, le filtre de Roberts est approprié.

Dans le cas le plus général, les bords correspondent à des changements de texture. Il peut donc être indiqué, dans des cas réellement complexes, d'appliquer l'extraction de bords non pas sur l'intensité lumineuse, mais plutôt sur une image élaborée, formée de descripteurs de textures calculés par ailleurs.

Typiquement, l'extraction de bords est suivie d'un seuillage qui élimine les bords peu marqués (donc peu fiables) et qui met en évidence les bords jugés significatifs.



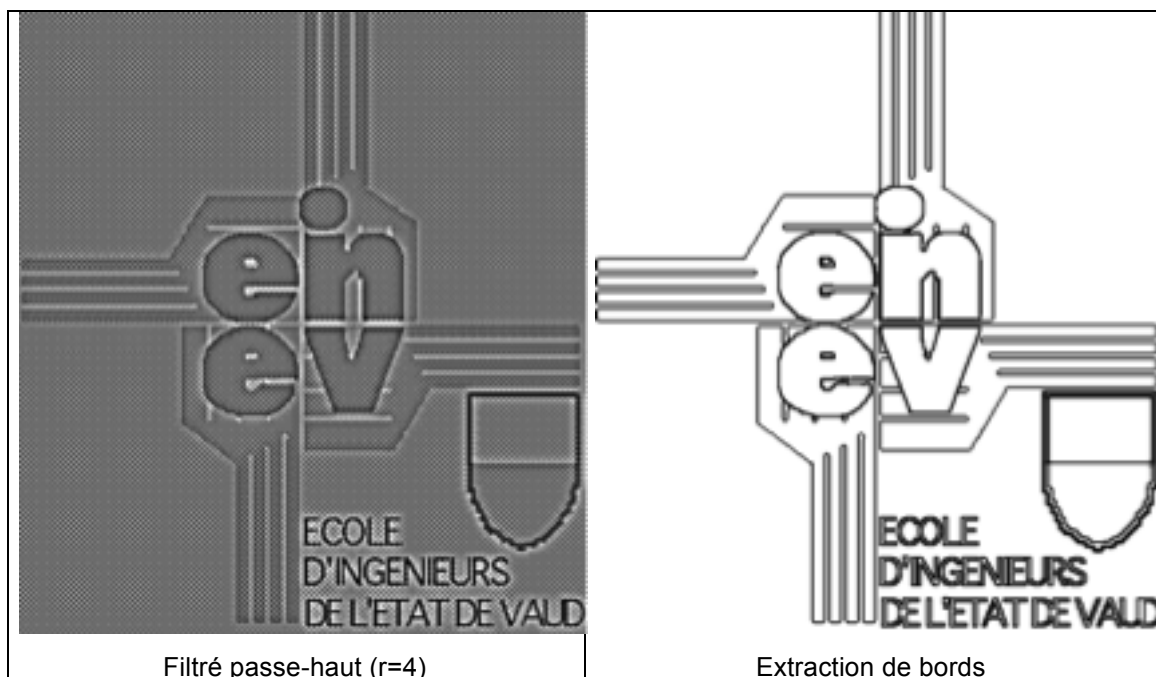


Fig. 3.7 Exemples d'images filtrées

Filtres adaptés

Il est souvent tentant de faire un filtre dont les caractéristiques correspondent exactement aux éléments d'images qui nous intéressent. Filtrer de cette manière donne des maxima ("point blanc") aux endroits dont les pixels correspondent à la forme d'intérêt.

Dans la version de base décrite ici, on parle en anglais de "template matching" (correspondance à un chablon). Il s'agit aussi en termes mathématiques d'une opération de "corrélation".

La théorie du traitement de signal amène une idée supplémentaire, qui consiste à éliminer les fréquences où le bruit est important par rapport au signal.

Filtrer avec une réponse impulsionnelle étendue (comprenant beaucoup de pixels) peut être très difficile en des temps brefs. L'augmentation du temps de calcul par rapport à la taille du filtre suit une loi en puissance 4, voire plus élevée. Il est donc souvent judicieux de détecter une partie seulement des formes intéressantes, en laissant à un traitement ultérieur, de plus haut niveau d'abstraction, le soin de déterminer si ces parties détectées forment un tout admissible.

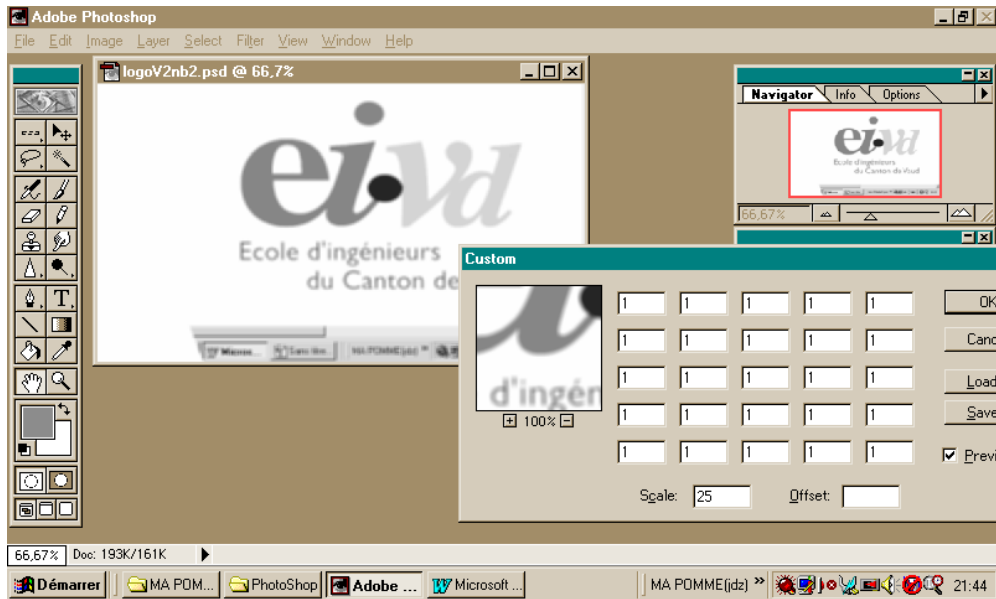


Fig. 3.7b Fenêtre de programmation d'un filtre linéaire sur Photoshop. Ici un filtrage passe-bas.

3.2.2 Filtres non-linéaires

Les filtres linéaires présentent l'avantage de s'appuyer sur un corps important d'outils théoriques. Il n'en va pas de même pour les filtres non-linéaires.

Dans une certaine mesure, les filtres non-linéaires peuvent s'interpréter comme des approximations linéaires, mais cela est limité. Le plus souvent, ils doivent être étudiés individuellement. Et les résultats sont le plus souvent l'aboutissement de démarches expérimentales.

Érosion et dilatation

Les opérations de dilatation et d'érosion sont utiles pour faire disparaître des pixels ou des taches. Les deux opérations sont complémentaires et sont souvent appliquées successivement.

Supposons que l'on veuille faire une érosion des blancs, codés chacun par des "1". La loi tient dès lors en une ligne:

Chaque point blanc dont l'un au moins des voisins est noir se remplace par un noir.

En définissant exclusivement deux pixels ayant une face commune comme voisins, l'opérateur d'érosion (des blancs) peut se décrire de la façon suivante:

$$p' = \text{Min} (p_2, p_4, p_6, p_8)$$

avec les conventions de noms (p' le pixel de sortie, p_2 le voisin nord, etc.) de la fig. 3.4. C'est en même temps l'opérateur de dilatation des noirs.

Les autres opérateurs, tels que la dilatation des blancs, s'obtiennent de façon similaire.

Bien que parfois ces opérations soient parfois tentées sur des images de gris, cela n'a guère de sens, et c'est au contraire dans les images binaires (booléennes) que ces opérations sont courantes.

Comme exemple d'application de ces opérateurs, considérons la suppression de traits et de taches noires de diamètre inférieur à 3 pixels. Par l'application répétée deux fois de l'opérateur d'érosion des noirs, ces éléments disparaissent. Les éléments noirs de taille plus grande demeurent, mais ils ont perdu deux "couches périphériques". On peut approximativement retrouver celles-ci en appliquant deux fois également, pour terminer le traitement, l'opérateur de dilatation des noirs.

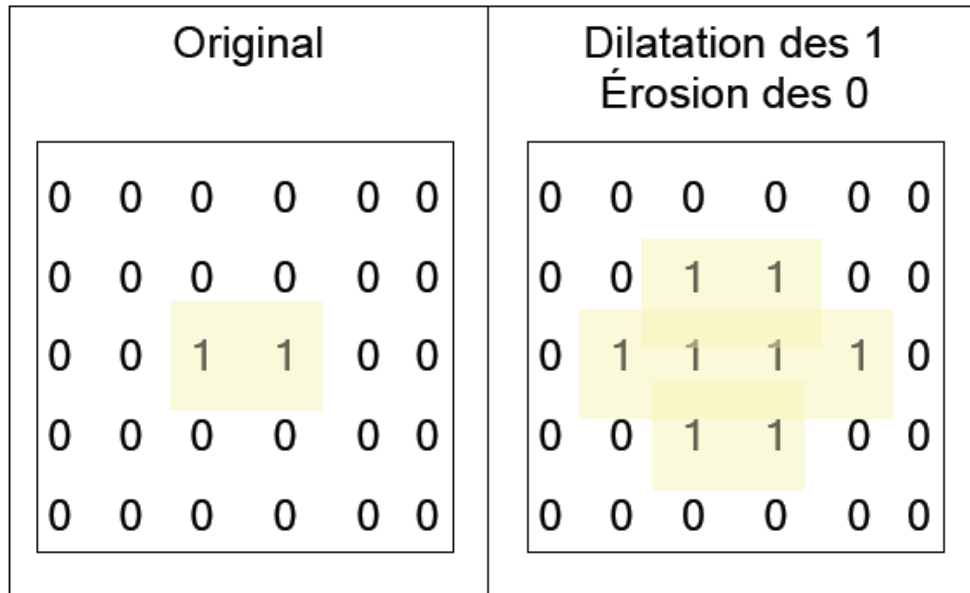


Fig. 3.7c Exemple d'érosion (région 0) et de dilatation (région 1)

Filtre médian

Le filtre médian se confond pratiquement avec le filtre passe-bas lorsque le noyau est de grande taille. C'est pour les petites grandeurs qu'il s'en distingue de façon utile.

La médiane d'un groupe d'échantillons se trouve par une opération de tri. Si les échantillons sont organisés de façon croissante, la médiane c'est la valeur de l'échantillon se trouvant au centre de la distribution. Dans le cas d'échantillons booléens (par exemple blanc et noir), l'opérateur médian se confond avec l'opérateur majorité.

Dans un noyau de taille 3x3, et avec les notations de la figure 3.4, le filtre médian peut se définir ainsi:

$$(e1.. e9):= \text{Tri}(p, p1 .. p8)$$

$$p'= e5$$

où, comme en langage Pascal, l'opérateur ':=' symbolise l'assignation.

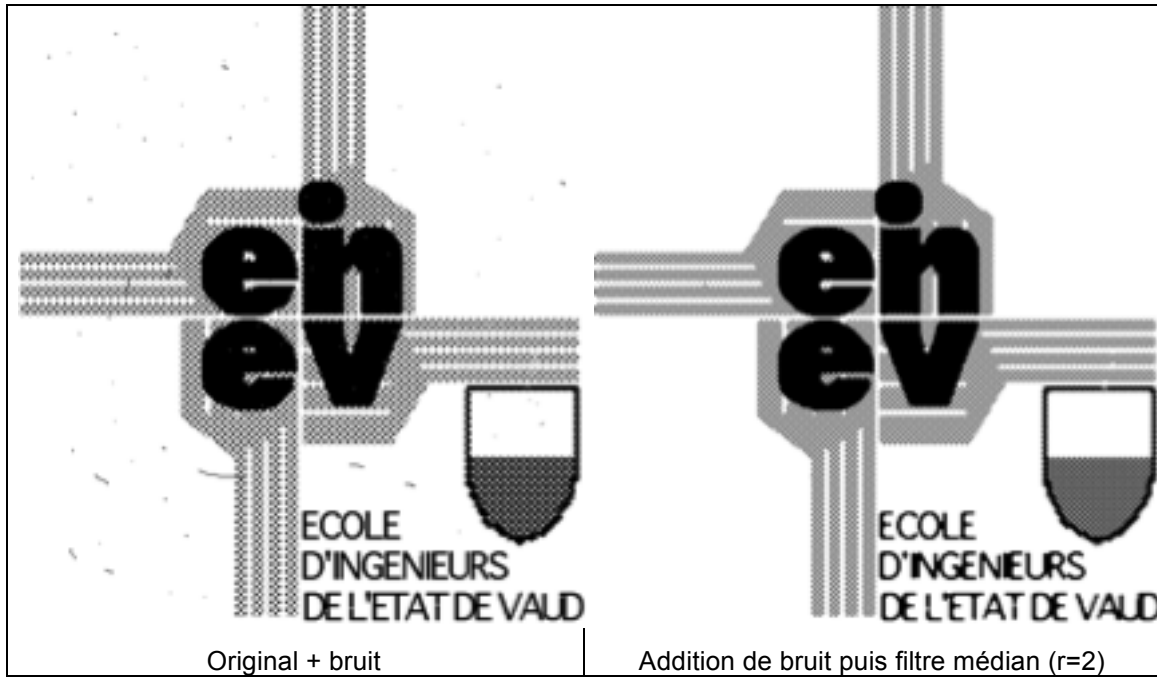


Fig. 3.8 Exemple d'application du filtre médian: les petits défauts sont éliminés, mais les coins sont légèrement arrondis, et parfois le défaut subsiste.

Squelettisation

Les images à nombreux niveaux de gris conduisent à des bords d'épaisseur variable, tantôt très large et parfois mince ou même inexistants.

Souvent c'est, du point de vue sémantique, la présence d'un bord (on dit aussi "contour") qu'il faut détecter et son épaisseur n'est guère utile. On ne peut dans ce cas simplement utiliser l'opérateur d'érosion vu ci-dessus. Car celui-ci, tout en amincissant valablement les segments épais détruirait de façon inacceptable des segments minces.

On définit donc des opérateurs de squelettisation dont la propriété essentielle est de limiter l'opération d'érosion aux seuls éléments non critiques pour garantir la continuité du bord.

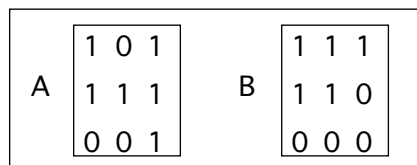


Fig. 3.9 Squelettisation. Le point central de la fenêtre gauche ne peut être supprimé car il est nécessaire pour relier deux groupes disjoints de voisins. A droite par contre, le pixel central peut être converti à "0".

Comme très souvent jusqu'ici, le traitement peut se faire en se basant sur le voisinage 3x3 de chaque pixel.

Morphologie mathématique

Les opérations de type érosion et dilatation ont reçu un fondement théorique sous le titre de "morphologie mathématique". Elles y ont été généralisées et enrichies de techniques diverses (notions d'"ouverture" et de "fermeture"), particulièrement adaptées aux images booléennes. Ceci apporte en particulier un substitut parfois efficace à l'opération de template matching introduite au § *Filtres adaptés*.

La notion fondamentale introduite est celle d'opérations déterminées par un "élément structurant". L'élément structurant est une petite image (groupe de pixels valant "1") que l'on peut à choix interpréter de deux façons équivalentes. De façon intuitive, on peut la considérer comme un chablon que l'on déplace sur l'image à traiter. De façon formelle, on l'interprète comme un ensemble de consignes de translations pour l'image sur laquelle elle agit.

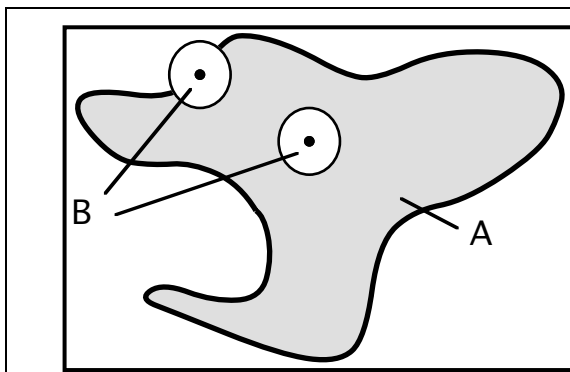


Fig. 3.10 Un élément structurant (B) est représenté en deux positions. Dans l'une, il est entièrement inscrit dans l'objet A. Dans la deuxième, ce n'est pas le cas.

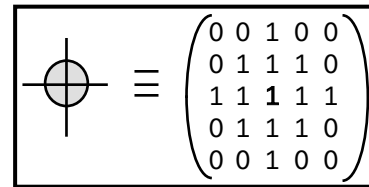


Fig. 3.11 Élément structurant; sous forme graphique simplifiée (gauche) et sous forme de matrice de pixels booléens (droite).

L'érosion, que l'on a vu plus haut dans son acception simple habituelle et redéfinie de manière plus générale en morphologie mathématique. On parle ici de l'érosion d'une image par un élément structurant (B). Etudions ce qui se passe pour un objet (par ex. A) de l'image. De manière intuitive, il s'agit alors de balayer l'image et de ne retenir que les pixels où le centre de B se trouve lorsque celui-ci est complètement inscrit dans l'objet A. De façon formelle, l'objet résultant est obtenu en faisant l'opération logique "ET" sur l'ensemble de toutes les translations de A définies par B (cf. fig. ci-après).

De même, la dilatation prend en morphologie mathématique une signification spéciale. On parle ici de la dilatation d'une image par un élément structurant (B). Voyons de façon détaillée ce qui se passe pour un objet (par ex. A) de l'image. De manière intuitive, il s'agit alors de balayer l'image et de retenir tous les pixels où le centre de B se trouve lorsque celui-ci touche au moins partiellement l'objet A. De façon formelle, l'objet résultant est obtenu en faisant l'opération logique "OU" sur l'ensemble de toutes les translations de A définies par B (cf. fig. ci-après).

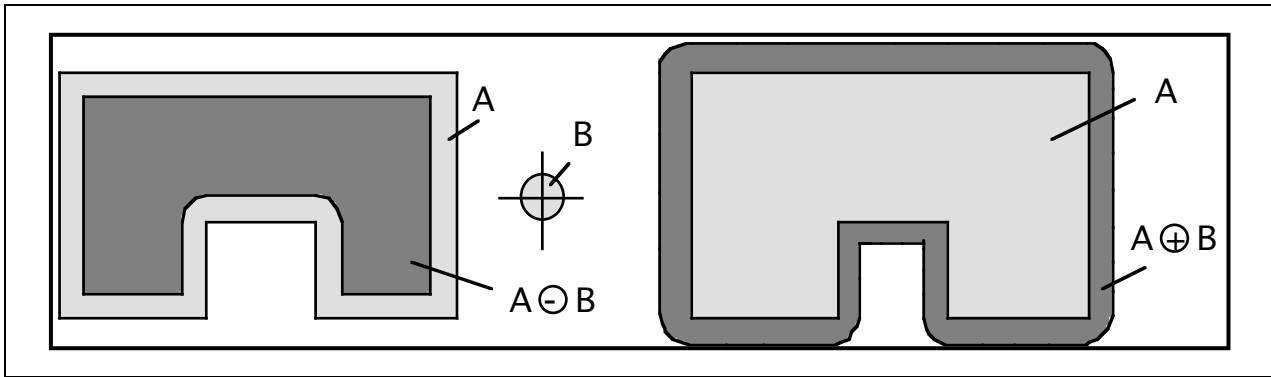


Fig. 3.12 Érosion de l'objet A par l'élément structurant B (gauche) et dilatation du même objet (droite)

De façon originale, les concepts d'ouverture et de fermeture ont été introduit en morphologie mathématique. L'ouverture (d'une image par un élément structurant), c'est l'application successive d'une opération d'érosion puis de dilatation (par ce même élément structurant).

Inversement, la fermeture (d'une image par un élément structurant), c'est l'application successive d'une opération de dilatation puis d'érosion (par ce même élément structurant).

Le mérite principal de la théorie dite de "morphologie mathématique", due pour l'essentiel à Jean Serra, de Paris, est de permettre la description compacte de variantes multiples possibles par applications successives des opérateurs traditionnels, simples, d'érosion et de dilatation. Cependant, il est clair que la charge de calcul, au niveau des pixels individuels, n'est guère affectée par le formalisme utilisé pour la description. La charge reste généralement très lourde.

3.2.3 Méthodes à relaxation

Il arrive que des caractéristiques globales, c'est-à-dire estimées à l'échelle d'une image complète ne corresponde pas très bien à certains voisinages. Ainsi par exemple, dans une scène où les objets apparaissent en clair par rapport au fond, un point à niveau de gris faible a une forte probabilité d'appartenance au fond. Mais s'il est entouré de points très clairs, manifestement appartenant à un objets, on peut considérer que sa chance d'appartenir à l'objet est plus grande.

Les méthodes à relaxation travaillent donc itérativement, les résultats d'une itération servant de base à la suivante. En général, pour être utiles, ces méthodes convergent après un certain nombre d'itérations vers une situation stable, jugée optimale (cf. Postaire 87).

Bien qu'intéressantes d'un certain point de vue théorique, les méthodes à relaxation se heurtent toutefois au problème des temps d'exécution, ce qui limite leur impact en pratique.

3.2.4 Espace des changements d'échelle (scale-space)

Une manière de limiter la quantité d'information à traiter à un moment donné consiste à n'analyser qu'une partie de la bande passante en fréquence spatiale. En d'autres termes, les variations rapides du

niveau de gris sont par exemple négligées au profit des variations lentes. Dans ce dernier cas, la théorie autorise un rééchantillonnage à cadence faible.

Un autre avantage de cette méthode de traitement à multirésolution, provient du fait que certains bords marqués pourront être détectés facilement et avec une bonne précision sur les images de basses fréquences. Aux fréquences plus élevées, on pourra éventuellement trouver d'autres bords, là où les "trous" de la première analyse demandent à être comblés. La représentation simultanée à plusieurs niveaux des données bidimensionnelles est parfois appelée la représentation *pyramidale*.

Parmi les défauts de la méthode, il faut noter la nécessité de traiter plusieurs images d'une même scène, chacune à des résolutions différentes, ce qui complique considérablement l'analyse. D'autant que des conflits peuvent apparaître entre les informations extraites des différentes images, qu'il faut bien résoudre.

Une bonne présentation de ce type d'approche est faite dans le numéro de sept. 90, de la revue PAMI (IEEE).

3.2.5 Conversion de texture

Lorsque l'analyse s'étend au delà d'un seul pixel, pour couvrir un voisinage plus ou moins grand, il peut apparaître des "régularités". Lorsque l'on a une région où le niveau de gris ou la couleur varient, mais où, néanmoins, on a la perception subjective de quelque chose de régulier, on parle de texture. Quelques exemples: la texture "sable" a des variations rapides dans toutes les directions, la texture "tissu" est caractérisée par des lignes horizontales et verticales.

En pratique, lorsque les niveaux de gris bruts d'une région doivent être analysés, il est très difficile de définir des lois de transformation qui fasse mathématiquement apparaître des régularités, c'est-à-dire une texture. On cherchera des constantes par exemple dans l'énergie (somme du carré des pixels), ou dans les coefficients de Fourier... Mais c'est un domaine encore très peu défriché.

Le voisinage de taille minimale permettant de caractériser une texture est appelé un texel (texture element).

3.2.6 Labellisation

La labellisation est intéressante pour segmenter une image (cf. §4.3 Labeling) et peut se faire essentiellement sur la base d'une fenêtre-opérateur (un « noyau ») de 3 pixels :

- pixel analysé
- pixel précédent sur la ligne et
- pixel précédent sur la colonne.

3.3 Opérations globales ou par régions

Certaines opérations se font sur des domaines relativement grands, pouvant aller jusqu'à couvrir toute l'image. A ce propos, on traite ci-dessous des opérations portant sur des fenêtres, puis de certaines techniques de compression d'images.

3.3.1 Opérations sur des fenêtres

Le travail par fenêtres est devenu populaire sur les ordinateurs avec l'arrivée des MacIntosh, et, plus récemment, avec l'avènement de logiciels tels Windows ou encore celui de la norme "Terminal X".

Il ne s'agit pas ici des mêmes concepts, car il n'est guère important, en analyse d'image, de travailler avec plusieurs fenêtres simultanées, et éventuellement superposées. Le problème est pourtant complexe aussi mais d'une autre façon. Les opérations principales sont ici celles de la gestion des bords de l'image (qu'il faut déplacer de cas en cas), de la rotation d'images, ainsi que des changements de résolution avec perte minimale d'information.

L'objectif des opérations sur fenêtre est double. Il s'agit d'une part de faciliter le travail de programmation de l'utilisateur. Il est aussi de permettre des traitements rapides par des opérations sélectives (taille et résolution minimales des fenêtres, avec changements dynamiques).

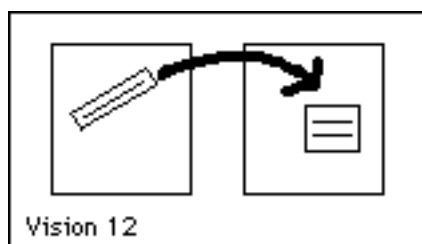


Fig. 3.13 Opérations sur fenêtres

Clipping et extrapolation

Lorsqu'une fenêtre est définie sur une image, la gestion des bords fait apparaître deux types de situations. Ou bien les bords de la fenêtre sont à l'intérieur de l'image, et la découpe se fait alors sans problème.

Ou alors, les bords de la fenêtre tombent, tout ou en partie, à l'extérieur de l'image, et se pose alors le problème de l'extrapolation de celle-ci afin de définir dans son entier le contenu de la fenêtre.

La méthode la plus simple consiste à remplir le fond par un niveau de gris constant (noir, blanc, ou moyen...). Mais une meilleure façon consiste à extrapoler les bords connus de l'image, tout au moins par un maintien (extrapolation d'ordre zéro). Une telle méthode réduira par exemple le risque d'apparition de faux contours aux limites des zones extrapolées.

La nécessité de couper une image, ou au contraire de l'extrapoler apparaît souvent lorsqu'on fait tourner une image.

Rotation

Il est souvent utile de tourner une image, afin de simplifier l'algorithme ultérieur de traitement. Par exemple, grâce à une rotation appropriée, l'accès aux pixels pourra par la suite se faire le long d'une colonne, ou d'une ligne.

L'opération de rotation comprend en fait deux étapes. D'une part, il s'agit pour chaque pixel de l'image d'arrivée, $p_s(l_s, c_s)$, de trouver le pixel correspondant dans l'image de départ, $p_e(l_e, c_e)$. De plus, ce pixel n'est généralement pas connu, et il s'agit de l'estimer en fonction de ses voisins.

Il est utile de travailler (balayer l'image) par rapport aux pixels de *sortie*, afin d'éviter que certains calculs se fassent à double, et surtout qu'il subsiste des pixels non définis, à cause des erreurs d'arrondis.

$$\begin{pmatrix} l_e \\ c_e \\ 1 \end{pmatrix} = \begin{pmatrix} \text{MATRICE DE} \\ \text{TRANSFORMATION} \\ \text{HOMOGENE} \end{pmatrix} * \begin{pmatrix} l_s \\ c_s \\ 1 \end{pmatrix}$$

L'interpolation peut se faire d'ordre zéro (reprise du voisin le plus proche):

$$p_s(l_s, c_s) = p_e(\text{Arrondi}(l_e), \text{Arrondi}(c_e))$$

Mieux encore, d'ordre 1, c'est-à-dire par interpolation linéaire à partir des 4 plus proches voisins:

$$p_s(l_s, c_s) = (c_+ - c_e) * (p_e(l_-, c_-) + p_e(l_+, c_-)) + (c_e - c_-) * ((l_+ - l_e) * p_e(l_-, c_+) + (l_e - l_-) * p_e(l_+, c_+))$$

où $l_- = \text{Entier}(l_e)$, $l_+ = \text{Entier}(l_e) + 1$, $c_- = \text{Entier}(c_e)$, et $c_+ = \text{Entier}(c_e) + 1$.

Rééchantillonnage (zoom)

Il arrive souvent que la résolution doive varier, soit pour des raisons d'intelligibilité pour l'observateur humain, soit pour optimiser les vitesses de traitement sur ordinateurs.

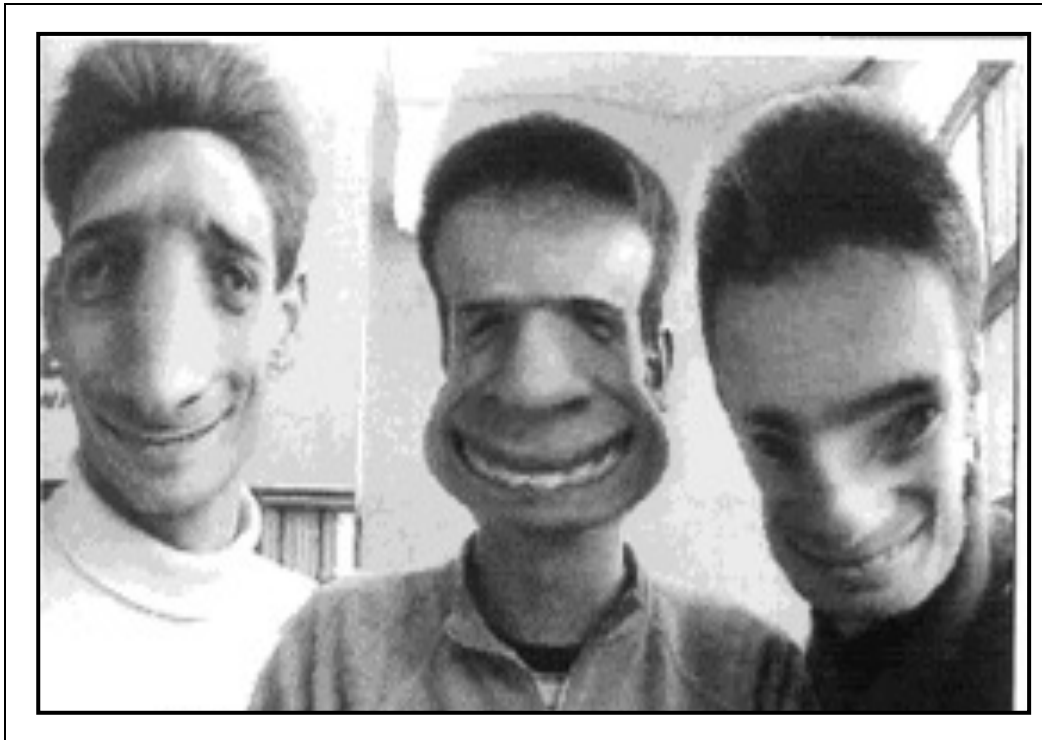
La solution la plus simple et la plus courante consiste simplement à sauter des échantillons lorsqu'on réduit, et à en répéter lorsqu'on agrandit (interpolation d'ordre zéro).

Mais il est parfois nécessaire de faire mieux: avant de réduire la fréquence d'échantillonnage, il faut s'assurer par un éventuel filtrage que le théorème de l'échantillonnage est respecté. De même en agrandissement, les résultats seront bien meilleurs si une fonction d'interpolation meilleure que celle d'ordre zéro est employée (ordre 1, comme au § précédent, ou autre). Une amélioration est aussi possible en filtrant passe-bas l'image agrandie.

Autres transformations spatiales

Il existe tout une classe d'applications où les transformations ne portent pas directement sur l'intensité lumineuse ou la couleur des pixels, mais bien plutôt sur leur position. En fait, la rotation d'images présentée ci-dessus relève déjà de ces techniques. Voici quelques autres cas de transformations spatiales:

- Compensation des distorsions géométriques présentes à l'acquisition d'images (imperfections des caméras à tube, projection sur une surface non plane, effets de perspective ou de parallaxe, etc.).
- Réorganisation du champ visuel pour simplifier l'adressage ultérieur. Par exemple le système Cognex permet d'exécuter très rapidement une transformation polaire: un cercle devient ainsi un segment de droite, et, de façon similaire, une couronne devient une bande rectiligne.
- Déformations "esthétiques" telles que projection sur une sphère, effet de vagues, ou encore déformations "élastiques" telles qu'utilisées pour les caricatures (cf. par ex. le logiciel Photostyler).



Traitement d'image avec opération sur les coordonnées spatiales ("Pinch")

De façon similaire, des modifications spatiales peuvent être opérées automatiquement sur une séquence d'image, afin, tout en respectant quelques points d'ancrage imposés par l'utilisateur, de faire une transition progressive entre image de départ et image d'arrivée (transformations morphologiques).

3.3.2 Techniques de compression

Depuis longtemps, pour la transmission de fac-similé, au aujourd'hui de fax, on a compacté les images en tenant compte qu'il y avait principalement 2 niveaux (noir, blanc) et une très forte prédominance des blancs. Le principe consistait alors généralement à transmettre non pas successivement l'état de chaque pixel, mais plutôt, le long des lignes successives, la longueur des plages blanches et/ou noires ("Run-length encoding", cf. §4.2 *Changement de représentation*).

Les besoins accrus de transmission et de stockage d'images ont conduit au développement de diverses techniques de compression/décompression des images couleurs ou à niveaux de gris. En particulier, avec l'avènement du multimédia, où des *flux* d'images sont traités, la nécessité de compacter temporairement les données est devenue évidente.

Il existe des méthodes ne perdant aucune information, mais elles ne permettent que des taux de compression faibles (facteur de compression allant jusqu'à 3). Pour faire mieux en compression, on doit admettre des pertes. Dans la mesure du possible, les pertes sont focalisées sur des composantes auxquelles l'oeil humain est peu sensible (en particulier: l'oeil ne perçoit pas les nuances d'intensité dans les zones à variations d'intensité rapides, et réciproquement, sa résolution spatiale n'est pas grande, là où les variations d'intensité sont lentes).

Parmi les méthodes standards de compression d'image, on citera celles du MPEG et du JPEG. Dans les deux cas, l'utilisateur peut choisir le compromis qu'il désire entre maintien de la qualité et performances en compression. En étant extrêmement tolérant quant à la qualité, on peut aller jusqu'à compresser environ 50 x la taille des données originales.

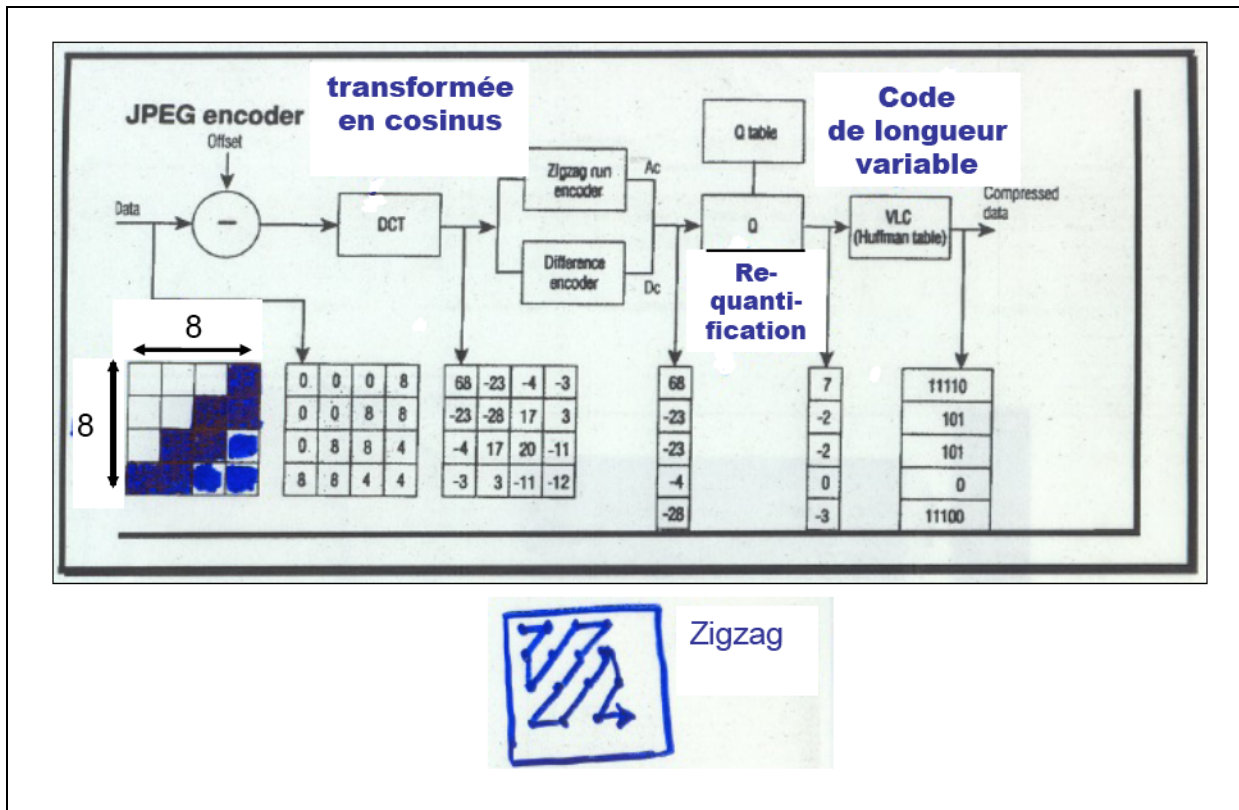


Fig. 3.14 Schéma-bloc du codeur JPEG (voir texte).

Dans la norme JPEG, l'image est traitée par blocs. Dans chacun d'eux, la transformée en cosinus (similaire à la transformée de Fourier) est calculée, puis on retient les coefficients de poids élevés, que l'on quantifie à peu de niveaux et que l'on code ensuite de façon à compacter l'information (code de Huffman).

Huffman table	
Symbol	Code
0	0
-1	100
-2	101
-3	110
-4	11100
+1	11101
+7	11110

Fig. 3.15 Table de codage pour compression selon la méthode de Huffman.

Dans le principe, la méthode de Huffman travaille comme suit: les valeurs peu courantes sont représentées par un symbole long, alors que les valeurs fréquentes sont représentées par un symbole très court. En moyenne, on peut donc se rapprocher du minimum correspondant au contenu effectif en information (cf. §1.2 *Capteurs et information*).

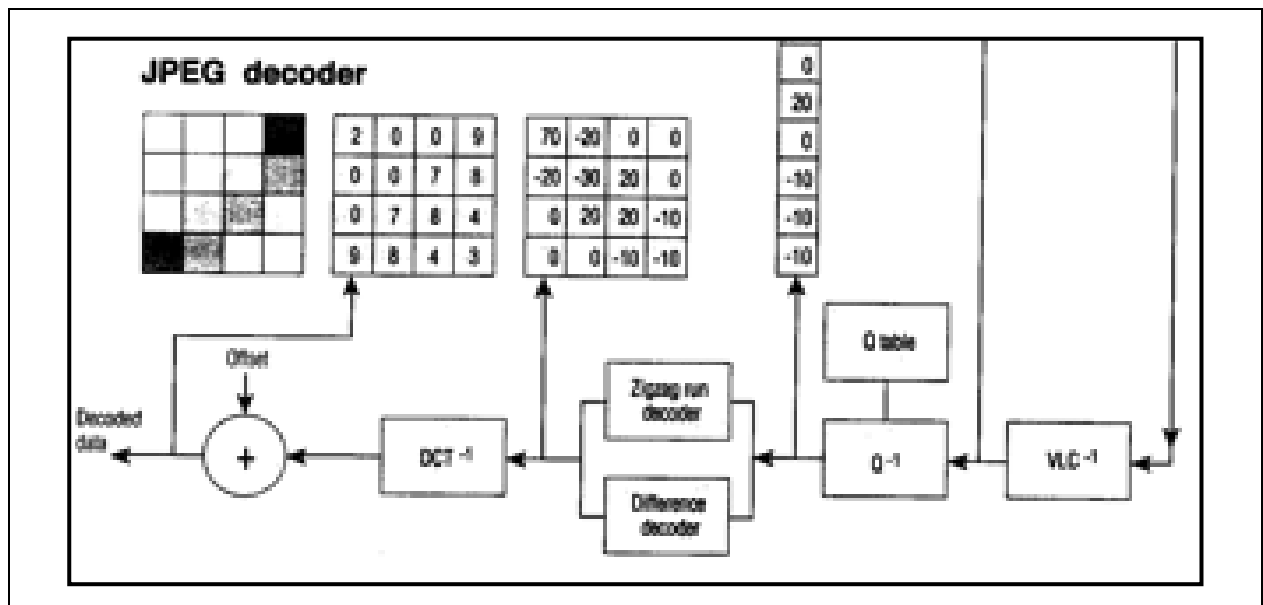


Fig. 3.16 Décodeur JPEG. Les opérations se font, au décodage, dans l'ordre inverse de ce qui se fait au codage.

A titre d'exemples, les figures 3.13. à 3.19 présente le principe du codage/décodage d'images selon deux standards actuels, JPEG et CCITT.

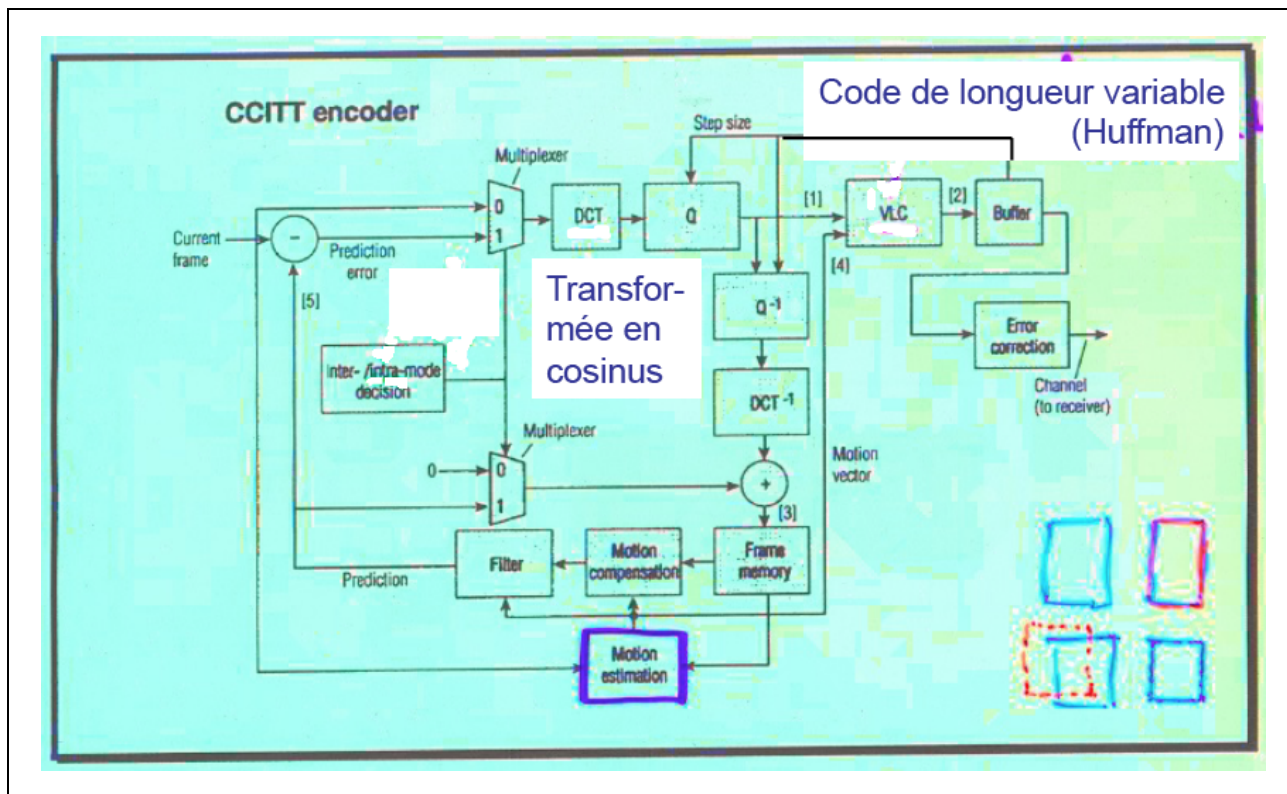


Fig. 3.17 Schéma-bloc du codeur CCITT (voir fig. 3.18 pour l'explication des abréviations).

Dans le codeur du CCITT que l'on présente ici il se trouve, en plus des éléments vus pour JPEG (transformée en cosinus, quantification, code de Huffman,...), une détection des mouvements de la scènes au cours du temps (notion de flux d'images). Pour éviter les dérives, le système redécote les éléments de sortie et prend ainsi en compte les images que le décodeur va recevoir pour calculer les différences entre images successives.

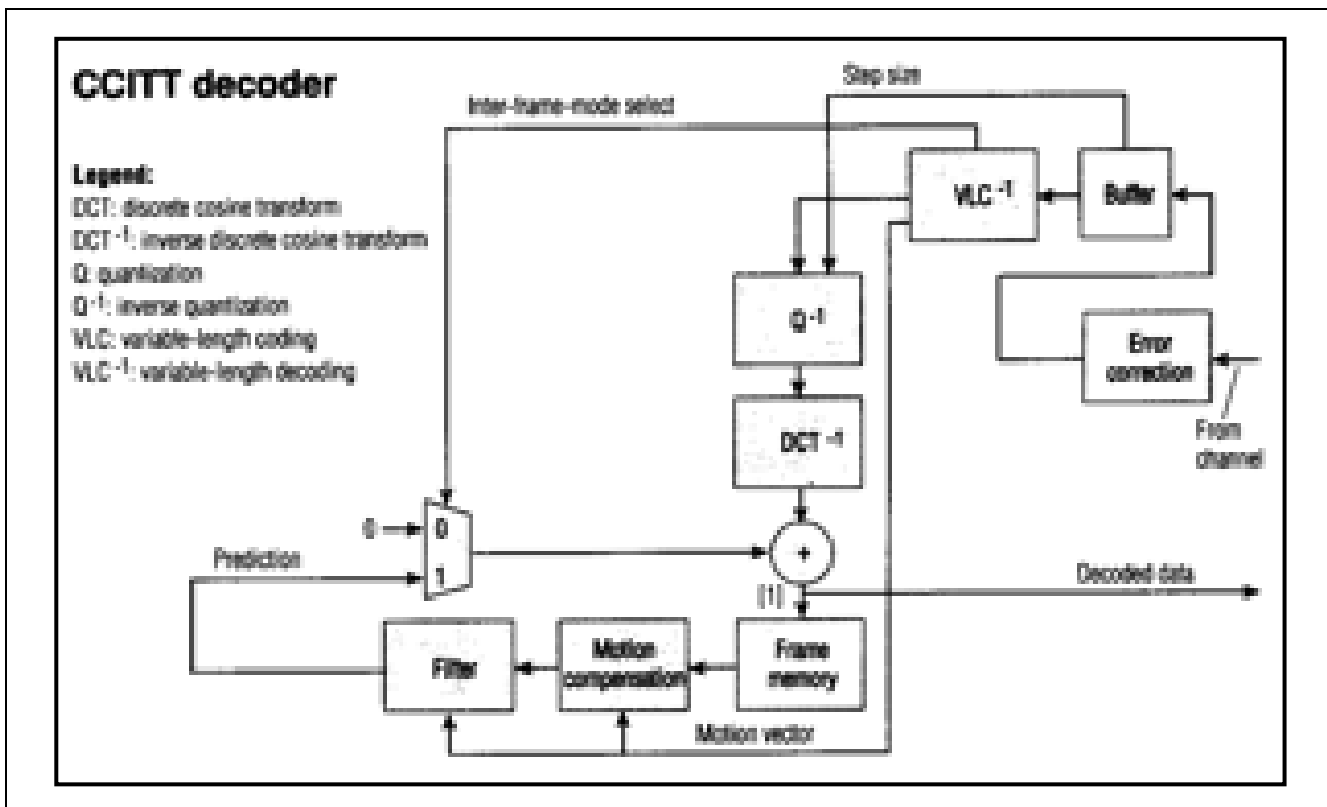


Fig. 3.18 Schéma-bloc du décodeur CCITT. Complémentaire au codeur présenté en fig. 3.11, le décodeur est en fait beaucoup plus simple.

Compression de séquences d'images: Références MPEG:

- Publications ISO, Genève, 022 749 01 11,
 - No 10'918 (2 parties, en préparation), et
 - No 11'172 (3 parties, 1993). 1: Séquences animées, jusqu'à 1.5 Mbyte/s, 2: Vidéo, 3: audio
- pages 428-430 dans M. Kunt, G. Granlund et M. Kocher, " Traitement numérique des images", Presses Polytechniques Romandes, 1993, pp. 562 .
- Image and video processing : 3-4 February 1993, San Jose, California / Majid Rabbani, M. Ibrahim Sezan, A. Murat Tekalp, chairs/eds. Bellingham, Washington : SPIE, cop. 1993 IX, 281 S. : Ill. ; 28 cm (Proceedings / SPIE - the International Society for Optical Engineering ; vol. 1903) "Sponsored by IS&T - The Society for Imaging Science and Technology ... et al." "The Conference on Video and Image Processing ..." ISBN 0-8194-1136-1 ETH-BIB P 713 011: 1903

3.4 Détection de mouvement

Pour une région d'intérêt, on détecte par corrélation ou méthode des différences (cf. §Comparaison de régions) la région correspondante (c'est-à-dire la plus ressemblante) dans l'image suivante (notion de flux d'image).

Dans le principe, la vitesse est alors estimée en divisant la distance (position relative) d'une région par rapport à l'autre, divisée par le temps séparant la prise des deux images.

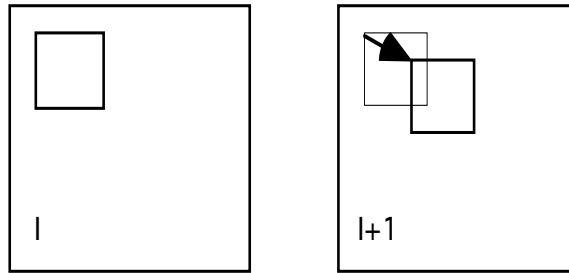


Fig. 3.19 Le vecteur indique le déplacement effectué par une région carrée (caractérisée par l'intensité de ses pixels) entre les images I et I+1 d'une séquence.

3.5 Transformée en cosinus

La transformée en cosinus 2D d'une image $x(k,l)$ est donnée par l'équation suivante:

$$X(m,n) = \frac{4C(m)C(n)}{N^2} \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} x(k,l) \cos \frac{(2k+1)m\pi}{2N} \cos \frac{(2l+1)n\pi}{2N}$$

où $m,n= 0, \dots, N-1$, et les coefficients $C(m)$, $C(n)$ valent 1 (sauf si m ou $n = 0$, auquel cas ils valent $1/\sqrt{2}$).

La transformée inverse est donnée par l'équation suivante:

$$x(k,l) = \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} C(m)C(n)X(m,n) \cos \frac{(2k+1)m\pi}{2N} \cos \frac{(2l+1)n\pi}{2N}$$

Cette transformée s'opère sur des blocs 8x8 dans les normes JPEG et MPEG

3.6 Temps de traitement

Dans le cas des traitements d'image, les temps de transferts de l'information sont relativement faciles à évaluer, et constituent en général une part majeur du temps de traitement total. Au temps mis pour accéder aux pixels d'opérandes et pour écrire les pixels résultants, il s'ajoute en fait un temps de calcul souvent très court.

On a donc une borne inférieure souvent très significative en limitant l'estimation du temps de traitement à celui des temps de transfert.

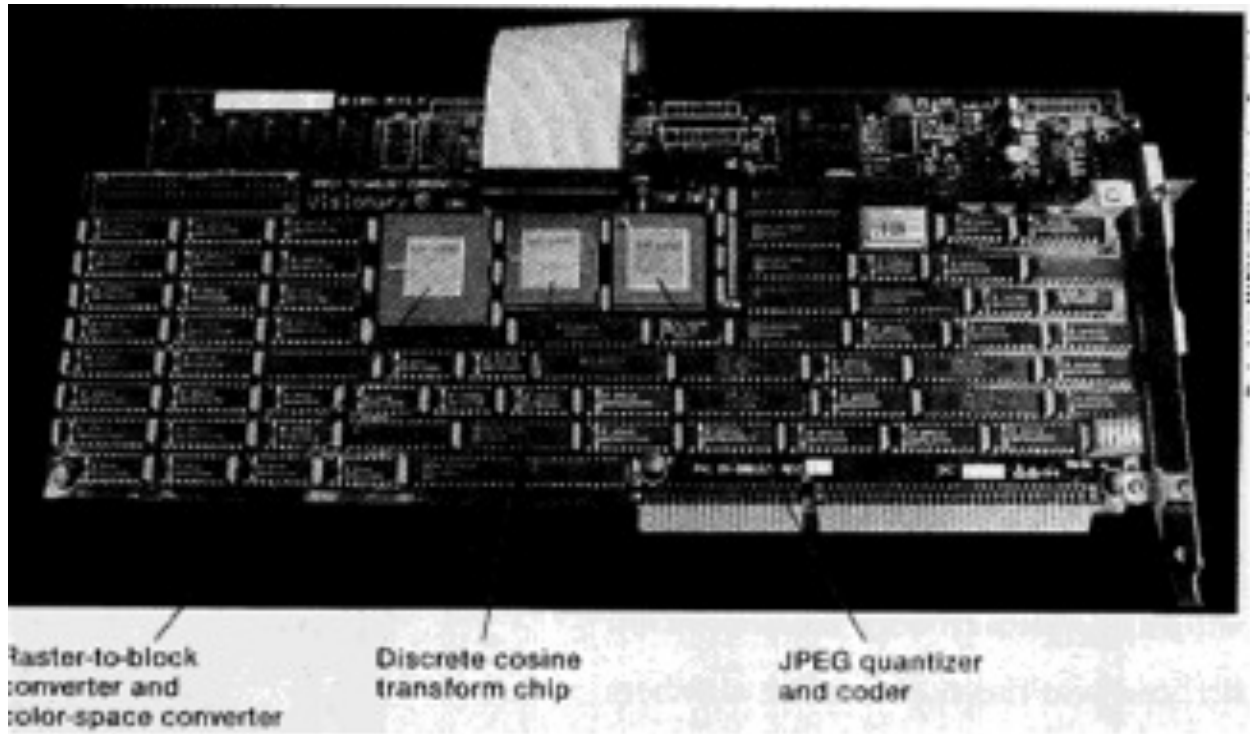


Fig. 3.20 Circuits pour le codage/décodage rapide (temps réel) d'images vidéo

4 ANALYSE DE SCENES

L'acquisition et le traitement bidimensionnel d'images font l'objet des chapitres précédents. Ces opérations préliminaires ne sont utiles qu'en vue de l'analyse ultérieure. C'est cette dernière phase que l'on discute ici.

Dans la plupart des cas où la vision par ordinateur est utile, on peut automatiser l'application de façon simple, fiable et économique en considérant l'image globale (image contenant par exemple 600x400 pixels) comme simplement un petit ensemble de photo détecteurs élémentaires virtuels (5 détecteurs ponctuels programmables en position ; ou barette de 100 photodiodes ; ou encore par exemple une couronne circulaire de 300 pixels). L'avantage est ici que la saisie d'image est bon marché, que le programmeur voit le contexte lors de la programmation et des tests, et que l'opérateur peut aussi juger de la situation pendant la production. Par ailleurs, au niveau du traitement informatisé, l'algorithme est simple à concevoir et à implémenter (par ex. vérification de la présence d'un trou, mesure de largeur d'une pièce, etc.)

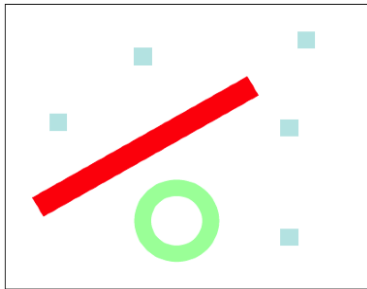


Fig. 4.0 Une image peut être simplement un ensemble de quelques pixels d'intérêt, d'une barre d'analyse, ou d'une zone circulaire. A droite, programme « Lireligne » du laboratoire, par exemple pour estimer le niveau de remplissage d'une bouteille d'eau

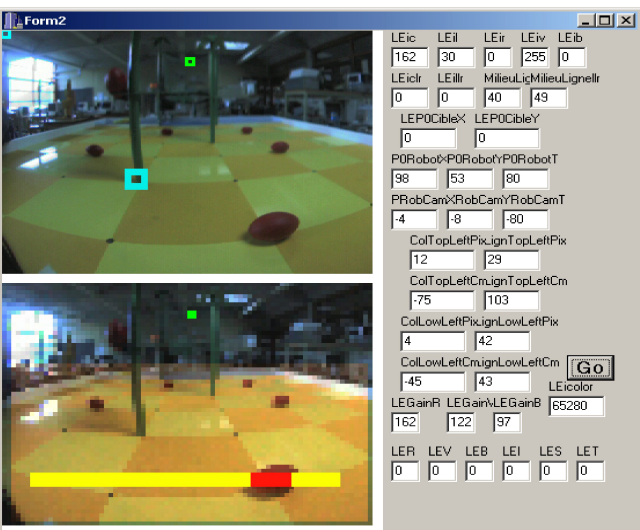
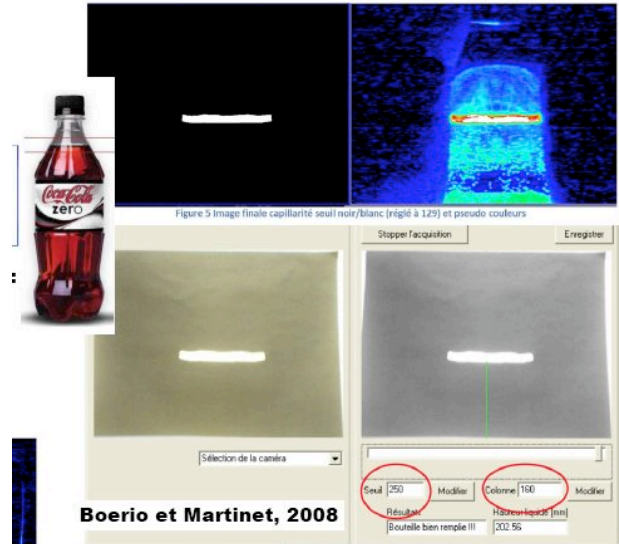


Fig. 4.0b Points et zones d'intérêt en contexte Piaget.

Dans des situations plus complexes, le traitement de pixels isolés ne suffit pas. Il faudra alors considérer des régions uniformes de l'image (« blobs ») et en extraire diverses caractéristiques primaires. Souvent, il est avantageux de changer de mode de représentation : par exemple, un contour peut consister en une suite de couples de coordonnées (ligne, colonne) ou se représenter par une suite de tangentes. Dans la plupart des cas industriels, la vision se limite au traitement d'images booléennes. Dans ce contexte, la segmentation qui permet d'isoler chaque objet vu prend un relief particulier, que nous discuterons avant de traiter des possibilités de reconnaissance des objets vus.

La section se termine par un paragraphe concernant les techniques d'analyse d'images permettant une estimation de la distance entre capteur et objets perçus.

4.1 Extraction de caractéristiques

Il n'est pas possible de gérer et de comprendre toute l'information contenue dans une image. Il s'agit donc ici d'en extraire un minimum, la seule qui suffise à satisfaire les besoins de l'application (par exemple: la position d'une pièce). Les données minimales ainsi extraites, ce sont les "caractéristiques": surface et périmètre d'une silhouette, position et orientation de pièces ou d'éléments tels que trous et coins.

Certaines caractéristiques sont utiles non pas directement pour l'analyse, mais permettent d'améliorer les étapes précédentes de traitement. C'est en particulier le cas de l'histogramme.

Histogramme

L'histogramme est essentiellement une fonction de probabilité non normalisée. La grandeur qui nous intéresse, par exemple le niveau de gris, est représentées par un certain nombre d'intervalles successifs, parfois appelés des boîtes. On répartit ensuite tous les échantillons dans ces boîtes en fonction de leur valeur. A la fin, l'histogramme informe donc sur le nombre total d'échantillons correspondant à chaque intervalle.

L'histogramme des niveaux de gris d'une image contient plusieurs informations utiles concernant son acquisition. Par exemple une image sous-éclairée aura un histogramme serré vers la gauche. Une bonne prise de vue, mettant en évidence des pièces blanches sur fond noir aura par exemple deux pics, nettement séparés par une vallée où il serait judicieux de placer le seuil de décision pour la classification des pixels (décision d'appartenance ou non aux objets).

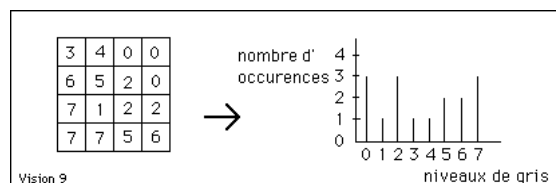


Fig. 4.1 L'histogramme est une fonction de probabilité non normalisée

Surface

La surface se définit facilement pour un objet unique dans une image booléenne. C'est la somme des pixels appartenant à l'objet.

$S = \sum_{\text{sur l'image}} (\text{poids}_i)$ avec $\text{poids}_i = 1$ si le pixel appartient à l'objet et 0 autrement.

Périmètre

Le périmètre d'un objet unique dans une image booléenne se définit de plusieurs façons. La plus simple indique que le périmètre est la somme de tous les points du bord de l'objet, c'est-à-dire de tous les points de l'objet qui ont un voisin dans le fond.

$P = \sum_{\text{sur l'image}} (\text{poids}_i)$ avec $\text{poids}_i = 1$ si le pixel appartient à l'objet et a un voisin dans le fond; 0 autrement.

Malheureusement cette définition ne permet pas d'estimation précise. Dans les cas défavorables, qui sont relativement courants avec les pièces manufacturées, les erreurs peuvent atteindre 20%, et ceci indépendamment de la finesse d'échantillonnage (résolution).

Pour une précision acceptable, le poids des éléments du bord doivent prendre des valeurs différentes de 1, qui dépendent de la forme locale du bord, suivant la Fig. 4.2.

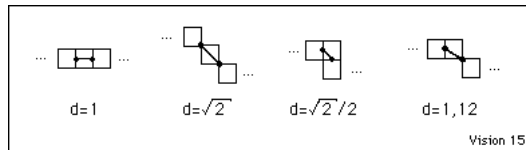


Fig. 4.2 Les éléments de périmètres ont une longueur effective qui dépend de leur voisinage.

Position

La position d'un objet unique dans une image booléenne, c'est-à-dire en général seuillée, est donnée par la moyenne des coordonnées des points appartenant à l'objet (Fig. 4.3).

$$x_g = \frac{1}{n} \sum_{k=1}^n x(k) \quad y_g = \frac{1}{n} \sum_{k=1}^n y(k)$$

Vision 16

Fig. 4.3 La position est souvent définie par le centre de gravité de la silhouette.

Les formules indiquées pour estimer la position correspondent à des moyennes. La précision du résultat est donc meilleure que celle des données. En général, l'estimation peut donc être plus précise que la résolution de l'image!

Orientation

L'angle de l'axe d'inertie minimum définit l'orientation de la pièce observée dans un plan perpendiculaire à l'axe du capteur (fig. 4.4 et 4.5). Cette méthode est attractive parce que les calculs mis en jeu sont simples; de plus l'évaluation peut se faire en cours de balayage.

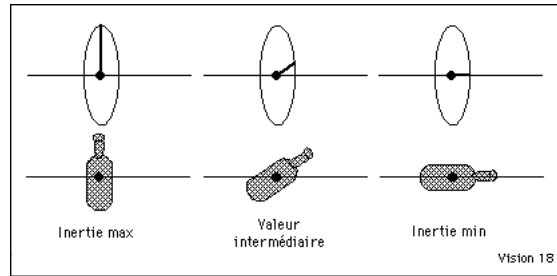


Fig. 4.4 Les moments d'inertie de l'image peuvent donner la position (cg) et l'orientation (axes de maximum ou de minimum d'inertie) de l'objet vu.

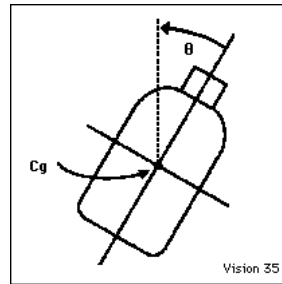


Fig. 4.5 Les moments d'inertie de l'image peuvent donner la position (cg) et l'orientation (axes de maximum ou de minimum d'inertie) de l'objet vu.

Mais elle a aussi deux aspects particulièrement défavorables: l'axe d'inertie minimum peut exister en nombre infini (c'est, p. ex. le cas pour un carré) et il demeure une ambiguïté de 180°, qui doit être levée par des heuristiques (Fig. 4.6).

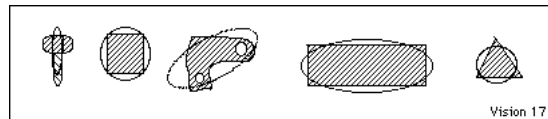


Fig. 4.6 L'ellipse d'inertie dégénère parfois en un cercle, et l'estimation d'orientation n'est pas fiable.

Les moments d'inertie de la silhouette peuvent se représenter sous la forme d'une matrice, M_I :

$$M_I = \begin{pmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{pmatrix}$$

dont les éléments sont définis par diverses sommes et produits calculés sur les coordonnées x_i, y_i des N pixels de la silhouette:

$$I_{11} = \sum_{i=1}^N x_i^2 - \left(\sum_{i=1}^N x_i \cdot x_{cg} \right) \text{ ou encore } I_{11} = \sum_{i=1}^N x_i^2 - N \cdot x_{cg}^2$$

$$I_{21} = I_{12} = \sum_{i=1}^N x_i \cdot y_i - \left(\sum_{i=1}^N y_i \cdot x_{cg} \right) \text{ ou } I_{21} = I_{12} = \sum_{i=1}^N x_i \cdot y_i - N \cdot y_{cg} \cdot x_{cg}$$

$$I_{22} = \sum_{i=1}^N y_i^2 - \left(\sum_{i=1}^N y_i \cdot y_{cg} \right) \text{ ou } I_{22} = \sum_{i=1}^N y_i^2 - N \cdot y_{cg}^2$$

Sur la base de ces moments, il est possible d'estimer l'orientation du grand axe d'inertie:

```

SI I12 ≠ 0 ALORS
    { Cas général }
    a := (I11 + I22)/2 ;
    b :=  $\frac{\sqrt{(I11 - I22)^2 + 4 \cdot I12^2}}{2}$ 
    Valeurs propres: λ1,2= a ± b;
    c := a+b - I11;
    α := Arc Tg  $\left(\frac{c}{I12}\right)$ 
SINON
    { Cas particuliers: ... }
    SI I22 > I11 ALORS
        α := π /2 { ... droite verticale, ou ... }
    SINON
        α := 0 ; { ... droite horizontale }
    FIN SI
FIN SI
    
```

Estimation de l'orientation par les moments d'inertie

La différence entre les valeurs propres λ₁ et λ₂ indique la fiabilité de l'estimation. Si elles sont très différentes, on aura donc nettement un grand et un petit axe. Dans le cas contraire, l'ellipse d'inertie tend vers un cercle, et l'estimation d'orientation opérée sur cette base devient très sensible aux perturbations.

L'orientation d'un objet peut encore être estimée par d'autres méthodes présentées plus loin: codage polaire, comparaison de courbures, corrélation. Elle peut aussi se baser sur la position relative de points particuliers (Fig. 4.7). Ces méthodes sont plus compliquées à mettre en oeuvre, mais elles sont d'emploi plus universel.

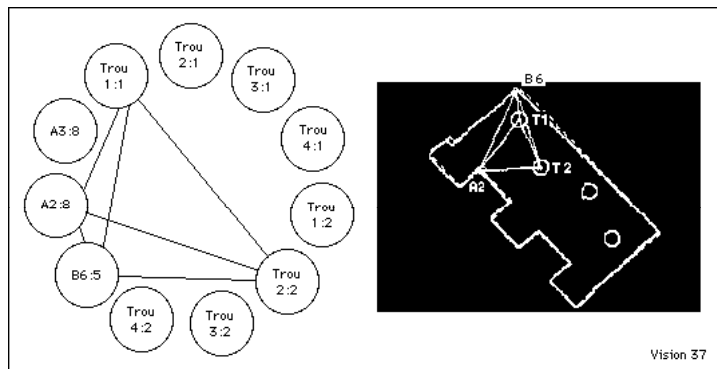


Fig. 4.7 La position relative de points particuliers (ici 2 trous, un coin convexe et un coin concave) permettent de définir la position et l'orientation d'un objet.

Points particuliers (prise, coins, etc.)

En plus des caractéristiques présentées ci-dessus, il arrive que l'on s'intéresse aux particularités de formes très locales: petits trous, coins, entailles, points de prise, etc.

La solution se trouve dans des contextes différents de cas en cas. Parfois, on se ramène à un identification de texture; ou ce sont des filtres adaptés que l'emploie; ou encore on analyse la courbure locale des contours; enfin, ce peut être l'analyse qui est reprise à une échelle plus fine: segmentation de l'image et interprétation des éléments segmentés.

Pour chacun de ces cas, le lecteur est renvoyé au paragraphe correspondant (texture, filtre adapté, etc.) situé plus haut dans le texte ou que l'on verra plus loin..

4.2 Changement de représentation

Il est utile de changer le mode de représentation de l'information pour plusieurs raisons. On le fait parfois pour réduire la quantité d'information, c'est-à-dire en extraire la partie particulièrement significative au vu de l'application. On le fait aussi souvent parce que certaines caractéristiques apparaissent clairement dans un mode de représentation alors qu'elles sont difficile à saisir sous une autre forme.

Le code de chaîne et le code de plage sont très utilisés pour comprimer l'information. Le codage polaire et l'extraction de courbure se révèlent précieux pour faciliter l'analyse. L'approximation polygonale est un cas intermédiaire qui apporte certains avantages tant sur le plan de la compression d'information que celui de la fonction de courbure.

Code de chaîne

Le code de chaîne s'utilise conjointement à la poursuite des contours. Lorsqu'on se déplace d'un élément de contour à son voisin immédiat, un nombre très restreint de directions sont possibles. Suivant la définition choisie, il y en a 4 ou 8. La méthode consiste donc à coder le contour par la succession des directions élémentaires. La longueur de la description est proportionnelle à la longueur des contours de l'image.

Ce code de chaîne est simple mais il ne permet guère l'analyse à cause de sa sensibilité aux variations d'orientation.

Code de plage

Le code de plage est une méthode très utilisée de compression d'information dans les images booléennes. Plutôt que de transmettre ou mémoriser chaque pixel successif, le principe consiste à indiquer la longueur des plages uniformes successives.

Par exemple la séquence "0001101111111000" peut être codée par les nombres "3,2,1,8,3". Typiquement, le code 0 correspond au noir, et le 1 au blanc.

Codage polaire

Le codage polaire pour un objet unique a le mérite de faire apparaître de façon évidente son orientation.

La Fig. 4.8 illustre le codage polaire. Dans une première étape, un contour C initialement représenté en fonction de x et y est reporté dans le plan r-θ. L'origine de ce plan est défini sur le centre de gravité de l'objet.

$$C: \{P_i(x,y)\} \rightarrow C: \{P_i(r,\theta)\}$$

avec $r = \sqrt{x^2+y^2}$ et $\theta = \text{Arc tg}(y/x)$.

Dans une deuxième étape, il est pratique de passer à une représentation univoque de r, r' en fonction de θ, et cela se fait par exemple en prenant la moyenne des valeurs multiples de r.

$$r'(\theta) = 1/n \cdot \sum_n r(\theta)$$

où n est le nombre de valeurs différentes de r au même angle. Ce nombre varie en fonction de θ. S'il est égal à zéro, r' n'est pas défini par la fonction et on peut le mettre arbitrairement à zéro.

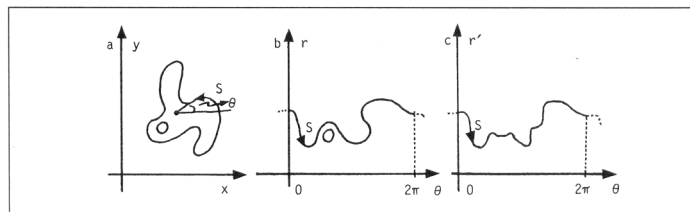


Fig. 4.8 Codage polaire d'une silhouette. Seuls les bords sont informatifs.

Ainsi, la représentation polaire est invariante à la translation de l'objet dans le champ de vue. De plus, une rotation de l'objet apparaît comme une simple *translation* dans la représentation.

Fonction de courbure

Passant en représentation polaire, on obtenait une fonction invariante aux translations de l'objet vu, ce qui simplifie l'analyse. Avec la fonction de courbure on va un pas plus loin: la représentation de contour par la fonction de courbure est invariante à tout déplacement, tant en translation qu'*en rotation*.

La courbure, c'est en principe la dérivée de la tangente au contour par rapport à l'abscisse curviligne (la longueur d'arc). On sait que l'opération de dérivation augmente généralement la sensibilité au bruit. Dans le cas de la courbure, cela est fortement le cas. Il faut donc impérativement filtrer le contour avant ou au moment de l'estimer.

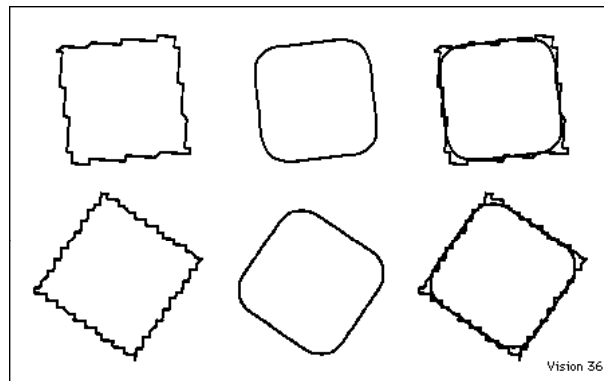


Fig. 4.9 Dans les images numériques, la quantification des contours conduit à un bruit local important. En conséquence la courbure ne peut s'estimer valablement qu'à partir d'un contour filtré. Ici c'est un filtre passe-bas simple (moyenneur) qui fournit la bonne solution.

Suivant le cas, cette définition de base peut conduire à l'utilisation de diverses formules. Mais toujours, la caractéristique de courbure présente les deux avantages suivants:

1. La courbure est définie très localement; elle s'applique donc indifféremment à des objets complètement visibles ou à de petits segments de contour.
2. La correspondance est évidente entre courbure et caractéristiques intuitivement très parlantes telles que coins, fentes, dents, etc.

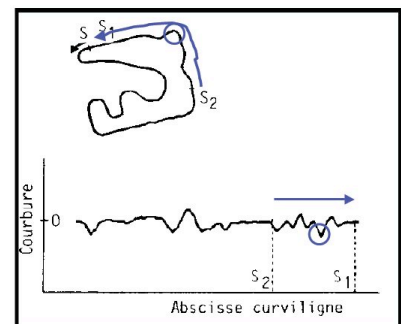


Fig. 4.10 Détermination de la pose d'une pièce au moyen de la courbure locale. La courbure est définie localement le long du contour. Elle est donc utile même si l'objet est partiellement caché.

Approximation polygonale

L'approximation polygonale d'un contour correspond à son remplacement par une suite de segments de droites. Cette approche permet un certain lissage et une certaine compression de l'information. Bien qu'assez courante parce qu'intuitivement évidente, cette approche n'est pas théoriquement bien fondée: Il s'agit en fait d'un sous-échantillonnage suivi d'une interpolation linéaire.

Lorsque l'interpolation se fait non par des droites mais par des courbes, il s'agit généralement de polynômes d'ordre supérieur à 1.

Un cas intéressant est celui des "splines". Dans ce cas, l'interpolation se fait par des polynômes. Mais en plus, des contraintes globales sur la courbe se substituent à l'information exacte de position des échantillons. Il s'agit donc plus d'une approximation que d'une pure interpolation.

Dans tous les cas, l'opération initiale de sous-échantillonnage conduit à des résultats peu prévisibles (phénomène de recouvrement).

Il vaut beaucoup mieux filtrer passe-bas les coordonnées successives du contour, ce qui, avantage supplémentaire, se fait de façon simple et rapide. Voici un exemple pour la coordonnée x (lignes):

$$x_s(i) = \frac{1}{N} \sum_{k=-N/2}^{N/2} x_e(i+k)$$

4.3 Segmentation

Bien des méthodes ne marchent que pour des objets uniques dans le champ de vue. Si l'on désire néanmoins utiliser ces méthodes lorsqu'il y a plusieurs objets, il faut segmenter l'image en morceaux.

Une approche convient bien aux images booléennes. Elle se base sur l'analyse de connectivité des pixels présents dans l'image. La deuxième technique présentée ici offre l'avantage de pouvoir s'appliquer aux images à niveaux de gris.

Analyse de connectivité

La méthode d'analyse de connectivité - également appelée parfois étiquetage ou "labeling" - convient bien aux images booléennes.

L'analyse se fait en un seul passage sur l'image, en considérant simultanément deux lignes voisines.

Trois situations typiques apparaissent:

- Détection d'un nouvel objet, et assignation d'une nouvelle étiquette
- Détection d'un pixel appartenant à un objet déjà partiellement vu et mise-à-jour de sa description, ou
- Détection d'un pixel appartenant à deux objets déjà partiellement vus, indiquant ainsi qu'en fait sous des étiquettes différentes on trouve deux segments d'un même objet.

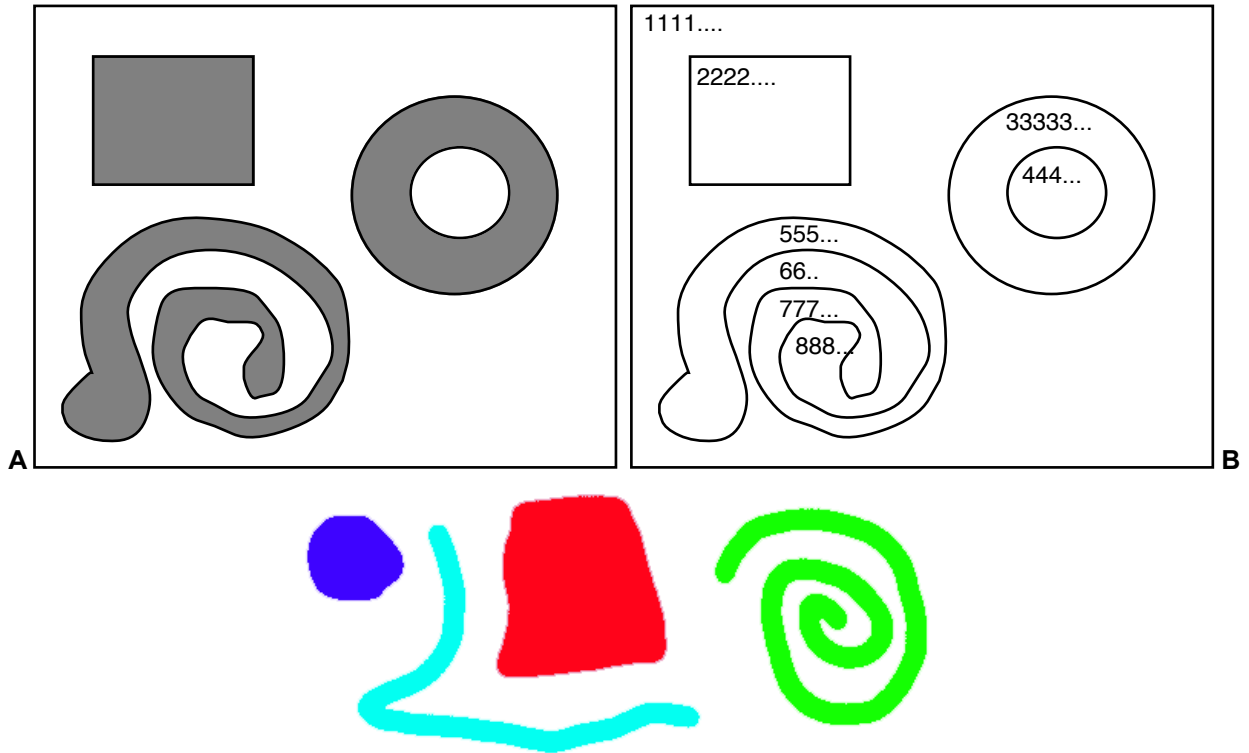


Fig. 4.11 Analyse de connexité. En analysant l'image booléenne A, on découvre différentes zones connexes, numérotées 1,2 etc. (B). En fait certaines zones se rejoignent (par ex. 5 et 7) et l'on découvre alors que l'une de ces deux étiquettes peut être remplacée par l'autre. Enfin, il est courant de coder par une couleur ou par un niveau de gris particulier, chacune des zones connexes rencontrées.

Poursuite de contours

Il est possible de segmenter une image en poursuivant le bord des objets visibles (voir fig. ci-dessous).

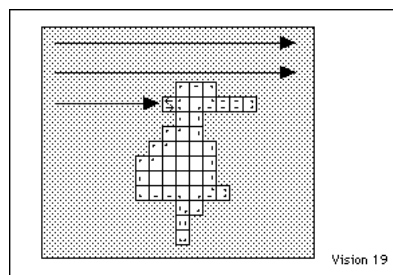


Fig. 4.11 Segmentation par poursuite du contour

Dans un premier temps, l'image est balayée de façon à détecter le bord d'un objet.

Ensuite, de façon systématique les voisins du point courant sont examinés de façon à se déplacer le long du bord (cf. fig. ci-dessous).

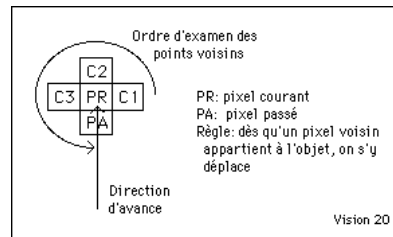


Fig. 4.12 Durant la poursuite du contour, les voisins du point courant sont analysés, et l'on se déplace le long du bord. Déplacement sur le premier candidat appartenant à l'objet.

La méthode peut se généraliser pour, à l'aide d'un deuxième pointeur, détecter le début et la fin de contours épais, encombrés de croisements ou au contraire s'arrêtant brutalement, comme cela se produit toujours durant l'analyse d'images à niveaux de gris.

4.4 Reconnaissance de "formes"

Les caractéristiques extraites des images servent parfois à estimer la position des objets vus. Mais il arrive aussi souvent qu'elles conduisent à une classification. Il s'agit alors de reconnaître un objet parmi d'autres, ou de vérifier que l'on ne dépasse pas une marge de tolérance préétablie.

Pour reconnaître un objet, plusieurs techniques s'utilisent, que l'on peut regrouper en deux catégories: les méthodes statistiques, et les structurales. Notons que "reconnaissance des formes" et la traduction répandue de l'anglais "pattern recognition". Alors qu'en français lorsqu'on parle de forme, cela évoque en premier lieu des caractéristiques visuelles, il est clair en anglais, que les "patterns" que l'on reconnaît peuvent être de toute nature: voix d'un individu, désordre particulier sur un électrocardiogramme, état mat ou brillant d'une surface, etc.

4.4.1 Méthodes statistiques

Les méthodes de reconnaissance de formes les plus courantes sont essentiellement statistiques. On discute ici les méthodes corrélatives, les distances dans l'espace des caractéristiques, ou d'autres méthodes encore.

Corrélation

La corrélation est une fonction mathématique permettant de comparer les signaux mono ou multidimensionnels. Elle est particulièrement intéressante lorsque les signaux subissent des perturbations aléatoires ("bruit") de type additionnel, car elle y est peu sensible.

$$R(k) = \sum_{i=1}^n x(i+k) \cdot y(i)$$

Dans l'équation, R est la fonction de corrélation. Si elle atteint des valeurs élevées, cela indique une bonne ressemblance entre les signaux x et y. De plus, la valeur k qui donne le plus grand résultat indique le décalage qui existe entre les deux signaux.

Pour les images, la fonction est souvent bidimensionnelle. Si l'on cherche par exemple à identifier un caractère d'imprimerie "A" dans une image "I", on glissera le caractère A le long des lignes et des colonnes et l'on retiendra les endroits où la fonction de corrélation donne de grands résultats:

$$R(\text{ligne}, \text{colonne}) = \sum_{l=1}^n \sum_{c=1}^n I(l + \text{ligne}, c + \text{colonne}) \cdot A(l, c)$$

où le caractère "A" est représenté par une petite fenêtre de $n \times n$ pixels.

On peut imaginer une corrélation de dimension 3 si le caractère a une orientation quelconque, ou même 4 si la taille est elle aussi inconnue. Ceci conduirait cependant à une charge de calcul importante.

Méthode des Différences

Similaire quant au but recherché, la fonction "Différence" est souvent préférée à la fonction de corrélation pour des raisons de calculs ou de compréhension intuitive:

$$D(\text{ligne}, \text{colonne}) = \sum_{l=1}^n \sum_{c=1}^n |I(l + \text{ligne}, c + \text{colonne}) - A(l, c)|$$

Alors que la correspondance la meilleure donne un maximum pour la corrélation (pic), on a par contre un minimum (creux) dans le cas de fonction "Différence".

Méthode HEIG-VD (iAi.LaRA) des Différences de couleurs pondérées selon saturation

A la HEIG-VD (iAi.LaRA), nous avons tôt pris en compte la saturation (représentation intensité-saturation-teinte, au lieu de RVB) pour considérer la teinte plutôt que l'intensité lorsque c'est possible, c'est-à-dire lorsque la saturation passe un seuil défini. D'où 9 couleurs (rouge vert bleu RVB, jaune cyan magenta JCM, noir gris blanc NGB).

De façon plus évoluée, les couleurs ont été reconnues avec une pondération dépendant de façon plus fine qu'au paragraphe précédent, variant linéairement entre deux seuils.

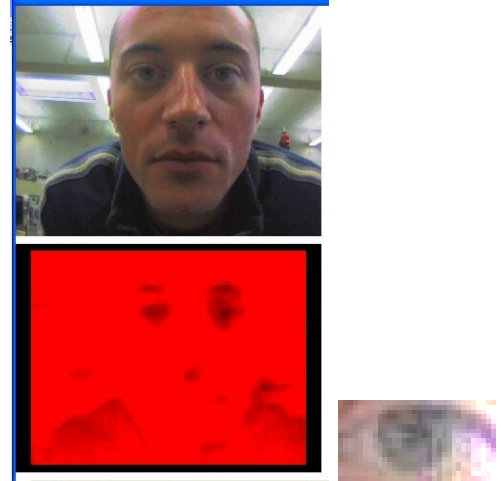
Enfin, l'idéal est une corrélation utilisant au niveau de la comparaison de chaque paire de pixels la différence (SbWCD) en valeur absolue de la couleur, au sens d'une pondération basée sur la saturation (sat), affectant la différence brute en teinte ($h_1 - h_2$, estimée cycliquement), et en intensité ($int_1 - int_2$):

$$SbWCD_e = \frac{w_{hue}(sat) \cdot (h1 - h2) + w_{int}(sat) \cdot (int1 - int2)}{w_{hue} + w_{int}}$$

, sans oublier la différence en saturation également (dContrast= sat1-sat2):

$$SbWCD = w_{contrast} \cdot dContrast + SbWCD_e$$

Fig. 4.12b Reconnaissance de l'œil gauche (tout à droite) dans l'image du haut par corrélation de ses couleurs, tenant compte d'une pondération des teintes et des intensités dépendant du niveau de saturation. L'intensité de rouge dans image du bas indique la différence locale intégrée sur une fenêtre de la taille de l'œil, successivement centrée en chaque point.



Distance dans l'espace des caractéristiques

Lorsqu'une seule caractéristique suffit pour distinguer deux formes - par exemple l'aire de la silhouette d'un objet- la comparaison est évidente. Mais lorsque plusieurs caractéristiques sont prises en compte, diverses manières de combiner leur apport sont possibles.

Il est habituel de représenter la forme à reconnaître comme un point dans l'espace mathématique engendré par ces caractéristiques. Jusqu'à la dimension 3, cela peut se traduire graphiquement (voir fig.). Mais au-delà, il faut simplement s'en tenir aux formules, en prenant garde que lorsque le nombre de caractéristiques croît (surface, longueur du périmètre, inertie, etc.) les probabilités qu'elles soient toutes indépendantes - et donc utiles - décroît.

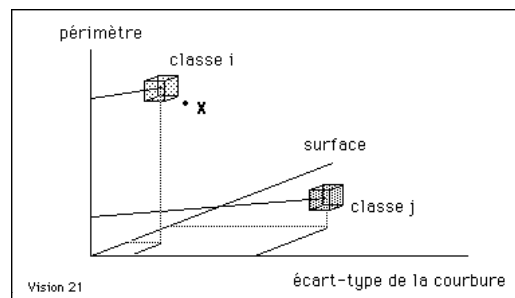
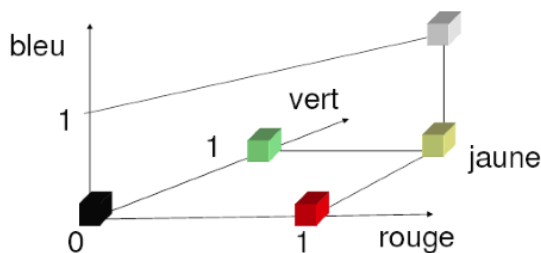


Fig. 4.13 Dans les deux exemples ci-dessus (couleurs à gauche, pièces mécaniques à droite), l'échantillon à reconnaître se traduit par un point dans l'espace des caractéristiques. Celui-ci est ici de dimension 3. Les petits cubes symbolisent la marge de tolérance que l'on admet par rapport aux formes-types.

Il est usuel d'estimer la différence entre une forme-échantillon et une forme-type en calculant la distance entre les deux points correspondants dans l'espace des caractéristiques.

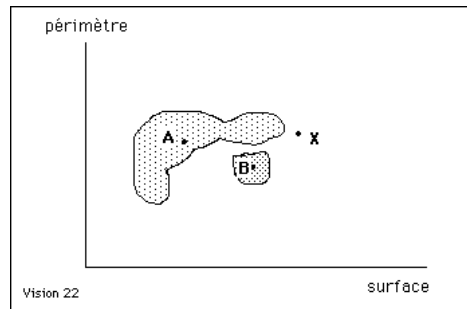


Fig. 4.13 Selon la règle du "plus proche voisin", il est convenu d'attribuer une forme à reconnaître, X, à la classe dont l'un des membres est le plus proche (ici: A), même si sa distance à la forme type est grande.

Autres méthodes

Il existe encore bien d'autres méthodes statistiques en reconnaissance des formes. Citons par ex. les notions de critère de Bayes, d'estimateurs, de mesures de convexité ... Mais ceci déborde notre cadre (voir par ex. Postaire).

4.4.2 Méthodes structurales

Les méthodes statistiques sont assez bien adaptées au cas simples (environnement fortement structuré, faible dimensionalité des variations). Mais lorsqu'on traite de cas plus complexes, tels que les objets naturels, ou les pièces industrielles flexibles, il faut recourir aux méthodes structurales.

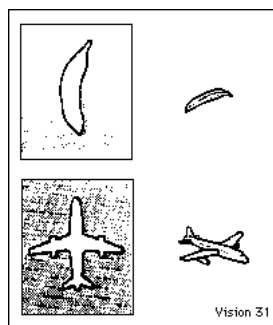


Fig. 4.14 Lorsque l'image varie de façon non-linéaire, les méthodes structurales sont généralement préférables aux méthodes statistiques. C'est le cas sur l'exemple, où l'on voit l'image varier par distorsions (formes naturelles) ou par effet de perspective.

On traitera ici surtout des arbres de décision, ainsi que des méthodes topologiques.

Arbres de décision

La reconnaissance d'une forme se fait parfois par une succession de décision individuelles. L'ensemble des possibilités de choix peut alors se représenter de façon arborescente: à chaque décision, c'est l'une parmi n branches qui est retenue, ce qui nous conduit sur un nouvel embranchement, et ainsi de suite.

Par rapport au cas précédent, la sensibilité au bruit de chaque comparaison élémentaire (c.-à-d. à chacune des décisions) est fortement augmentée. Par contre, la faible interdépendance entre décisions individuelles permet de s'accommoder à de grandes distorsions.

Si l'on désire reconnaître une pièce X parmi n modèles, on commencera par exemple par comparer sa plus grande dimension. Sur cette base, un ou plusieurs modèles seront écartés. C'est ensuite une autre caractéristique que l'on observera, et ainsi de suite on s'acheminera sur la bonne classe.

Méthodes topologiques

Un ensemble particulier de méthodes structurales est celui qui regroupe les méthodes topologiques.

Prenons l'exemple d'une pièce caractérisée par la position relative de détails locaux, tels que des trous et des coins. Avec les méthodes topologiques, on ne s'intéresse généralement plus de façon fine à connaître les distances entre ces caractéristiques, mais on se satisfait de la description de leur existence, ainsi que d'une description qualitative de leur relations mutuelles (en particulier, position relative).

Autres méthodes

Les méthodes structurales comptent encore bien d'autres techniques, telles que les approches "grammaticales", les graphes d'appartenance, ou encore les modèles entités-relations... Pour une étude plus poussée de ces notions, on se reportera par exemple à [Niemann].

4.5 Estimation de distance

La distance entre un objet et un capteur peut être estimée de nombreuses manières. Nous complétons ici, en ce qui concerne particulièrement la vision, les informations déjà données dans le § *Capteurs* de la partie 1. Nous voyons successivement les techniques du temps de vol, de la triangulation, avec en particulier la stéréoscopie, de la radiométrie, et de la "mise au net".

Durée de vol

La plus intuitive se base sur le temps de propagation d'une onde. Une onde acoustique est adéquate si l'objet n'est pas localisé avec une grande précision (cf. relation entre résolution et longueur d'onde). Cette méthode est utilisée avec succès en photographie amateur et peut être un complément utile au sens visuel des robots. Lorsque l'onde utilisée est électromagnétique (p. ex. rayon infrarouge ou lumière), une bonne résolution spatiale est liée à des temps de propagation extrêmement courts, de l'ordre de la nanoseconde. Néanmoins des capteurs apparaissent maintenant qui exploitent ce principe (par exemple la caméra Swiss Ranger SR3000 du CSEM, produite en 2006, a une résolution d'environ 200 lignes par 200 colonnes, et peut détecter des distances allant jusqu'à environ 7 mètres. Le capteur capte aussi l'intensité lumineuse.

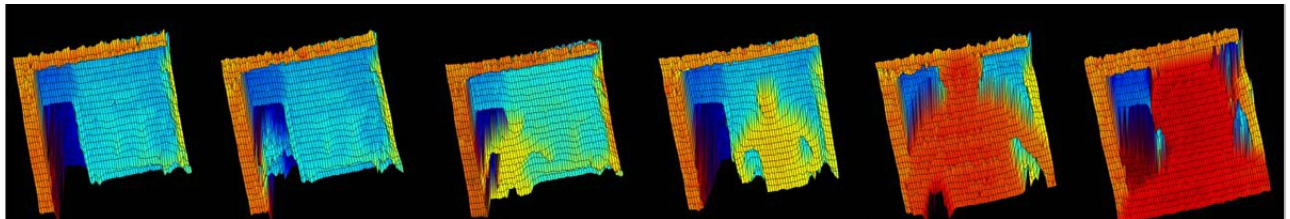
SwissRanger 3000 Camera Characteristics



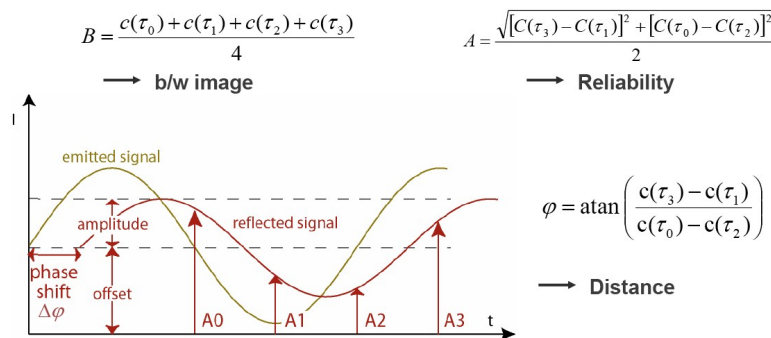
- Size
 - 50mm x 48mm x 67mm
- Weight
 - 175g
- Power Supply
 - 12 V @ 1 A max
- Interface: Default USB2.0
 - Read & Write registers
- QCIF sensor resolution (176x144)
- Background suppression
- Distance resolution: see Demo

csem centre suisse d'électronique et de microtechnique

3D-TOF SR-3000 | Image Sensing | Page 15



Phase-Measurement TOF Principle - Theory the 4-Tap Algorithm



csem centre suisse d'électronique et de microtechnique

3D-TOF SR-3000 | Image Sensing | Page 11

Triangulation

Les méthodes d'estimation de distance les plus courantes pour les objets proches sont basées sur la triangulation. On utilise un point lumineux qui balaie la scène. Si le point est visible, l'angle du rayon perçu sur le capteur, la distance entre le capteur et la source lumineuse et l'angle d'émission du rayon lumineux permettent de connaître la distance entre ce point et le capteur. En utilisant un laser et des filtres optiques appropriés, il est possible de faire des mesures rapides (10^4 points par seconde) en lumière ambiante même avec une puissance émise modeste de l'ordre du mW. La limitation de la méthode vient des zones «d'ombres» où le spot lumineux n'est pas perceptible sous l'angle correspondant au capteur visuel. Un certain gain de vitesse est possible si l'on projette simultanément

toute une ligne, voire toute une matrice de points au lieu de balayer l'objet à localiser par un seul point. Il faut toutefois noter que des ambiguïtés sont alors de plus en plus probables, entre points émis et points perçus.

Des travaux de recherches actuels (Prof. Kanade, à l'Université de Carnegie-Mellon) indiquent qu'il serait possible, avec des circuits intégrés spécialisés, de détecter à très grande vitesse (env. 100x par seconde) l'ensemble des distances séparant l'observateur de chaque point visible d'une scène. L'idée est qu'en chaque pixel, il se trouve un circuit capable de détecter l'instant où le signal reçu est maximal (détection de pic). De façon synchrone, un plan lumineux balaye la scène. Comme la méthode implique l'éclairage de la scène, elle ne peut pas s'utiliser pour des distances supérieures à quelques mètres.

Une méthode également basée sur la triangulation, mais qui se distingue des précédentes par l'absence de lumière spécialement projetée sur l'objet est très utilisée en photogrammétrie: c'est la vision stéréoscopique. Les difficultés de mise en œuvre de cette technique en robotique sont liées à plusieurs facteurs. D'une part, une corrélation d'images est requise, qui est gourmande en ressources de calcul. De plus, cette corrélation doit se faire de manière locale («par morceaux»), car chaque zone d'image corrélée doit correspondre à tout ou partie d'un seul objet. Enfin, on a des difficultés fondamentales:

- 1) éventualité de zones d'ombre où une zone perçue par un capteur ne l'est pas par l'autre;
- 2) nécessité d'avoir une texture suffisamment riche et aléatoire pour permettre la corrélation précise de toute zone visible;
- 3) sensibilité aux différences des deux capteurs utilisés en stéréoscopie et à l'éclairage dont la perception varie avec la position des capteurs.

Un argument positif provient de la disponibilité actuelle, à bas prix, de circuits intégrés spécialisés pour calculer une corrélation selon un axe. Ces circuits sont couramment utilisés en photographie.

Théorème du cosinus. En triangulation, on fait souvent usage du théorème du cosinus. Rappelons ici sa définition.

Chaque angle d'un triangle peut s'exprimer en fonction des trois côtés:

$$c^2 = a^2 + b^2 - 2 ab \cos \gamma$$

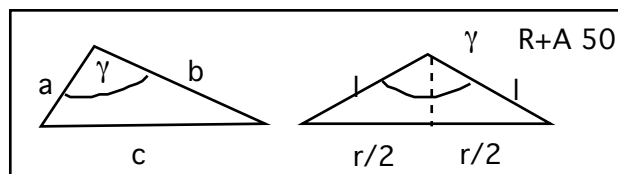


Fig. 2.6.5 Lorsque les côtés d'un triangle sont connus, on peut en calculer les angles

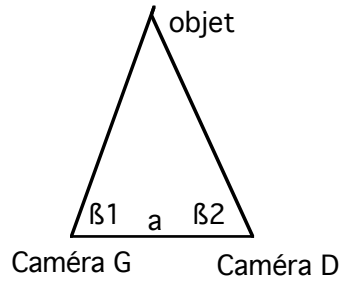
Si le triangle est isocèle, l'angle γ peut s'obtenir de façon plus simple encore:

$$\gamma = 2 \arcsin (r / 2 l)$$

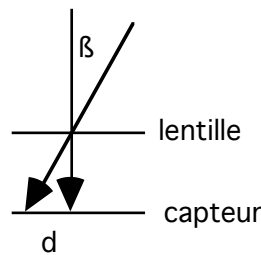
Même si un seul côté est connu, on peut calculer les deux autres pour autant que deux angles au moins soient connus. C'est cette dernière propriété qui est la plus utilisée dans l'estimation de distance.

Stéréoscopie

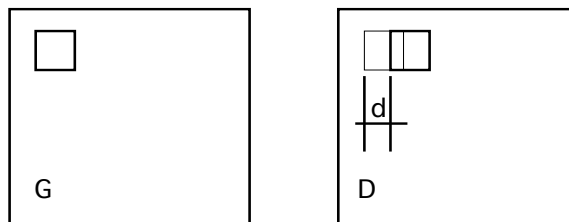
En stéréoscopie, on vise à l'aide de deux caméras un (élément d')objet dont on veut connaître la distance. D'abord, une comparaison est faite des deux images (cf. §*Comparaison de régions*), de façon à identifier les régions correspondantes. Puis on estime le décalage relatif de ces régions. Enfin on utilise la méthode de triangulation.



Les angles β_1 et β_2 varient en fonction de la distance entre objet et caméras.



Equivalence entre angle d'incidence et position (d) sur l'image



En stéréoscopie, un même objet apparaît dans les images gauche et droite avec un décalage relatif, d, qui est fonction de sa distance aux caméras.

Radiométrie

La radiométrie utilise le fait qu'un élément de surface d'objet réémet (la méthode est surtout intéressante pour des objets qui ne brillent pas, c'est-à-dire dont la surface diffuse la lumière) une partie de la lumière incidente non seulement proportionnelle au coefficient de réflectance, mais qui est également fonction de l'angle relatif entre la normale à l'élément de surface et les rayons de lumière incidents. Un capteur et trois sources lumineuses allumées tour à tour permettent ainsi de connaître l'orientation des surfaces

visibles, ce qui est lié à la dérivée de la distance entre capteur et objet. Outre la quantité des calculs nécessaires, la méthode a le défaut de n'être applicable qu'à la zone linéaire du capteur. Celle-ci est limitée à moins de deux décades. D'autre part la radiométrie est similaire à l'interférométrie en ce qu'elle ne donne pas de valeur absolue, même grossière, de la distance.

Mise au point de l'objectif

La distance peut en principe s'estimer également en "observant" la focale lorsque l'image perçue est nette. Ce point a déjà été discuté dans la partie 1 (§ Capteurs).

Ajoutons ici que le traitement d'images requis pour détecter que l'image est nette est basé sur des considérations relatives au spectre spatial de puissance.

Le flou d'une image correspond à son filtrage passe-bas. Lorsqu'on observe, en fonction des composantes en fréquences spatiales, la puissance de l'image perçue, on constate que sa bande passante augmente - et en conséquence la puissance totale - à mesure que la focale se rapproche de sa valeur optimale. Un test consiste dès lors à intégrer, dans la zone d'intérêt, l'intensité de chaque pixel élevée au carré.

La sensibilité augmente si l'on somme le carré de la différence entre l'amplitude de chaque pixel et le gris moyen de l'image. Elle augmente plus encore si l'on se concentre sur les fréquences élevées, en intégrant le carré de l'amplitude de chaque pixel par rapport à son *voisin*.

4.6 Problème général de l'estimation de pose

Au moment d'estimer la pose d'un objet, la distance de l'objet au capteur est, en général, approximativement connue. Ceci définit un facteur d'échelle entre l'objet dans l'espace et sa projection sur le capteur. Il sera pris égal à 1 dans cette discussion.

De façon informelle le terme pose décrit indifféremment, pour un objet, sa position, son orientation, ou une combinaison quelconque de ces deux grandeurs. De façon plus formelle, une pose est définie comme un point dans un espace à six dimensions, P , généré par toutes les translations et les orientations possibles d'un objet rigide. Un état, dans notre contexte, est défini comme l'une parmi plusieurs régions disjointes de l'espace P . Il est dit stable s'il est associé à une quantité d'énergie (localement) minimale. L'énergie est liée à la présence d'un champ (en pratique, la gravité ou un champ magnétique) et à des forces et moments transmis mécaniquement. Les problèmes d'estimation de pose sont caractérisés par la dimension de l'espace des poses possibles (EPP) correspondant. Par exemple, on définira comme un problème d'estimation de pose monodimensionnel, l'évaluation de la position d'une pièce qui ne peut bouger que le long d'un axe. On dira également que l'EPP est de dimension 1.

L'estimation de pose varie largement en complexité en fonction de la multiplicité des objets dans le champ de vue, de leur superposition éventuelle, de la rigidité des objets ou de limites mécaniques qui peuvent restreindre leur EPP. Dans les applications industrielles, il apparaît que les pièces sont généralement analysées une-à-une, reposant sur une surface plane.

Lorsqu'un objet se trouve isolé sur une surface plane, il peut généralement se décrire par sa silhouette. Dans ce contexte, l'EPP est tridimensionnel: deux translations et une rotation dans le plan de travail. Un bon contraste peut, en général, s'obtenir et des solutions utilisant une représentation binaire de l'intensité lumineuse sont bien établies. Dans ce cas simple, des produits commerciaux pour l'estimation de pose sont disponibles. Les techniques les plus puissantes sont basées sur un codage polaire (fig. 7), ou sur l'utilisation de caractéristiques locales particulières (coins, petits trous, etc.). Elles sont rapides (environ 0,1 s par objet) lorsqu'implantées sur du matériel spécialisé. Une autre classe de solutions est particulièrement populaire pour réalisation logicielle. Celle-ci regroupe les méthodes d'estimation de moments d'inertie de l'image. Si les pièces reposent sur un plan mais sont partiellement cachées (superposition) des méthodes basées sur l'utilisation de caractéristiques locales et, en particulier, de la courbure du contour sont possibles (fig. 9). Dans la mesure où les pièces ne sont visibles que partiellement, un risque d'erreur sur l'identification des pièces existe, qui est inévitable. Il n'existe pas de solution pratique pour les situations où l'EPP est supérieur à trois. Une difficulté fondamentale vient peut-être du fait que les capteurs visuels perçoivent une *surface* dans l'espace et non l'espace dans ses trois dimensions, comme en tomographie par exemple.

Le problème de base décrit peut aussi se compliquer énormément si la forme des pièces change, comme dans les opérations de forge, le démoulage ou lorsque les pièces ne sont pas rigides.

L'estimation de pose d'un objet n'est généralement pas indépendante de son identification. Si l'objet est perçu dans sa totalité (absence de recouvrement et d'ombre) des paramètres indépendants de la pose peuvent être définis, tels que la longueur du périmètre et la surface circonscrite par le périmètre (contour). Ces paramètres sont toutefois peu nombreux et modérément précis [14]. Ils ne permettent donc de différencier que quelques pièces de types très différents. Pour l'estimation de pose, par contre, un plus grand nombre de paramètres sont généralement requis. Ils peuvent alors aussi s'utiliser pour l'identification de pièces même similaires. Si une pièce n'est pas totalement visible, l'interdépendance est encore plus nette puisque c'est la position relative de détails identifiés (structures locales telles que coins, fentes ou trous) qui permettent l'estimation de pose. Dans ce contexte, la courbure du contour qui est une fonction invariante à la translation et à la rotation des objets est beaucoup utilisée.

Si un objet a une pose connue comme, par exemple, une pièce rigidement fixée sur son support, la tâche d'identification est alors considérée comme faisant partie du travail d'inspection.

4.7 Inspection

La tâche d'inspection se caractérise par différents traits dont les plus importants sont les suivants: pose de l'objet inspecté connue (au moins grossièrement), grande quantité d'information perçue par le capteur et, en conséquence, traitement hiérarchisé, analyse par régions et algorithmes heuristiques. L'orientation et la position connues a priori, cela permet de focaliser localement l'inspection. Ainsi une résolution poussée dans les quelques (petites) zones adéquates de l'image reste compatible avec un temps

d'analyse admissible. Si la pose de l'objet n'est pas connue, l'échantillonnage doit être uniforme pour toute l'image, et une étape d'estimation de pose précède l'inspection.

Il apparaît que les applications en inspection nécessitent généralement une haute résolution et un vaste champ de vue (en grandeur relative). Par exemple, une pièce de quelques cm doit être mesurée au 1/100 mm, des rayures de l'ordre du μm doivent être détectées sur un cristal de quelques mm ou encore des trous de quelques 1/10 mm doivent être reconnus sur des pistons de 10 à 20 cm de diamètre. Il existe des exemples similaires dans l'industrie textile, électronique (circuits imprimés) etc.

Vu la quantité d'information perçue par le capteur, il est vital, en inspection, de hiérarchiser le traitement. Ainsi, dans une application industrielle d'inspection d'étiquettes, une première phase estime l'écart entre la position effective de l'étiquette et sa position nominale. Pour cela, l'état d'un bord de l'image perçue est analysé. Pour la suite du traitement, l'image perçue est déplacée pour correspondre exactement à une image de référence interne. Dans une autre application, la lecture de quelques colonnes de l'image suffit à affiner l'estimation de pose d'une dent de scie en carbure de tungstène (fig. ci-dessous). Par la suite, une fraction seulement de l'image à 256 x 256 points est analysée pour détecter des défauts de bord (bords cassés); en cas de test positif, une dernière zone de l'image est traitée pour vérification de l'état de surface.

Étant donné la variété des tâches possibles en inspection et le temps d'analyse limité, il n'existe pas de solution générale. Il est nécessaire de définir de cas en cas la stratégie adéquate. La communication entre opérateur et système d'analyse doit donc être particulièrement soignée. On définit interactivement des zones d'intérêt dans l'image (fenêtres), petites et peu nombreuses. Pour chaque fenêtre, une résolution aussi grossière que possible doit être choisie, pour chacun des deux axes du capteur indépendamment. A l'intérieur des fenêtres, le traitement le plus adéquat doit aussi être défini. Ce sera au besoin une estimation de dimension, de surface, de niveau de réflectance, éventuellement une estimation de pose ou encore d'autres paramètres.

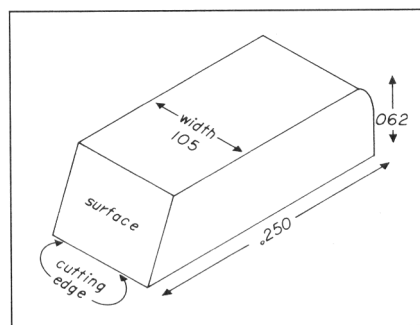


Fig. 4.15 Inspection d'une dent en carbure de tungstène. D'abord la largeur doit être vérifiée (dimension), puis le bord tranchant (cassures éventuelles) et, en cas de tests positifs, l'état de la surface de coupe.

4.8 Prise de pièces en vrac

Bien que l'on s'efforce de maintenir la position et l'orientation des pièces à travers les divers postes d'une usine, il arrive très souvent que le désordre s'installe. Parfois, pour des raisons technologiques (p. ex. pour l'opération de sablage), mais généralement pour des raisons économiques les pièces sont simplement mises en vrac dans des caisses. Les méthodes courantes pour l'utilisation de pièces ainsi entreposées se divisent principalement en deux classes: dans l'une, de la main d'œuvre prend les pièces à la main et en alimente les machines ou des magasins-tampons; dans l'autre, des bols vibreurs trient et orientent les pièces (fig. ci-dessous).

On ne reconnaît pas encore suffisamment l'acquisition de pièces en vrac comme une tâche importante pour les robots, et pourtant il le faudrait. Deux raisons à cela: d'une part, l'acquisition automatique de pièces apparaît souvent comme le dernier maillon à automatiser le long d'un processus de fabrication. Ainsi, dans plusieurs pays, on a développé des ateliers dits flexibles qui peuvent usiner et assembler des pièces manuellement préparées sur des



Fig. 4.16 Bol vibreur pour le positionnement de pièces stockées en vrac

D'autre part, des instituts de recherches ont prouvé la faisabilité de cette automatisation par des moyens robotisés [5, 6]. Pour être soluble, le problème de la prise de pièces en vrac doit se formuler sous une forme élémentaire: «Étant donné un organe de préhension, où dans un tas de pièces, est-il possible de prendre quelque chose?» Ce qui est fondamental, c'est d'identifier un point de prise. En conséquence, l'estimation d'orientation et l'identification de la pièce prise sont reportées à une phase ultérieure, après que la pièce acquise soit séparée des autres par le robot et que sa pose soit éventuellement mécaniquement limitée à quelques degrés d'incertitude [5].

Avec les méthodes publiées à ce jour, la probabilité d'acquisition d'une pièce (inconnue) d'un tas atteint souvent 90% après trois essais. Une des premières méthodes utilisées permet la détection de points de prise pour une ventouse (fig. ci-dessus). Elle se satisfait d'une représentation booléenne à 128 x 128 points par image traitée. A la suite d'un prétraitement qui segmente partiellement l'image et qui réduit la quantité de points à traiter, des zones uniformes, libres de bords sont détectées. Les plus grandes zones sont alors désignées comme points de prise possible (fig. ci-dessus). Des expériences ont donné d'excellents résultats pour des pièces aussi variées que des cylindres de tailles diverses (acier), des

pales de turbines pour avions à réaction (titane forgé), des boîtes de raccordement (aluminium) et des stylos feutres (plastique).

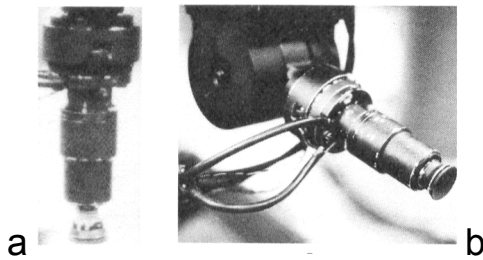


Fig. 4.17 Organe de préhension doté d'une ventouse. En (a), la ventouse peut s'accommoder à la surface de prise en laissant passivement varier son orientation dans un domaine de 120°. En (b), la ventouse est rétractée pour forcer la surface de prise dans un plan perpendiculaire à son axe. La main est équipée d'un capteur de contact .

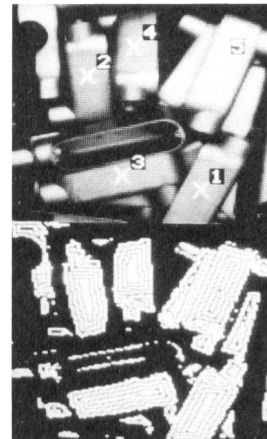


Fig. 4.18 Acquisition de pièces en vrac. A droite, cinq points de prise pour la main de la figure 3 ont été détectés et sont superposés à l'image numérisée. A gauche, un détail du traitement est illustré.

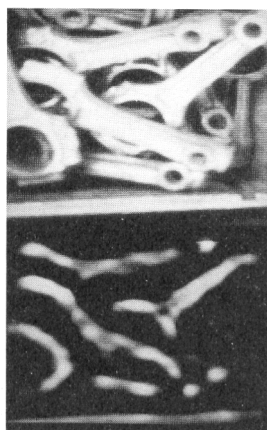


Fig. 4.19 Représentation à niveaux de gris multiples. A gauche, l'image numérisée de pièces en vrac (128 x 128 pixels à 6 bit). A droite, l'image après filtrage adapté, l'intensité indique en chaque point la probabilité d'y avoir un bon point de prise, pour la pince de la figure ci-dessous.

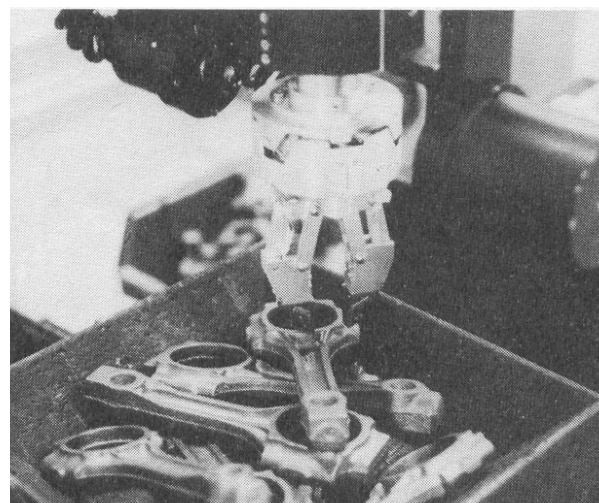


Fig. 4.20 Cette pince à deux doigts est l'un des préhenseurs utilisées pour la prise de pièces en vrac.

Pour une autre classe d'applications où la surface (en termes de réflectance, de texture ou de forme) des pièces ne conduit pas à un bon contraste entre pièces superposées, une représentation à niveaux de gris multiples est nécessaire. Dans ce domaine, une technique basée sur l'usage de filtres adaptés s'est révélée intéressante. Une zone de préhension typique doit d'abord se définir. Celle-ci dépend à la fois des pièces à manipuler et de l'organe de préhension. Elle se définit naturellement en termes

géométriques. Un modèle permet ensuite de passer de la représentation géométrique à une représentation lumineuse, telle que perçue sur les capteurs visuels. Ceci définit un filtre spatial (si possible récursif) qui s'applique numériquement à l'image. Les zones de préhension probable apparaissent alors comme des zones d'intensité élevée (fig. 4.19) et s'identifient donc par une simple détection de pics. Ainsi, par exemple, une pince à doigts parallèles (fig. 4.20) a des chances de prendre une pièce là où deux volumes vides entourent un volume occupé par un segment de pièce. Pour l'unité de vision artificielle, deux zones noires entourant une zone plus claire doivent se détecter.

5 PARTICULARITES DE LA VISION POUR ROBOTS

Les robots doivent, par définition, s'adapter à une grande variété de tâches. Mais jusqu'à présent, l'absence de capteurs extéroceptifs a présenté un obstacle majeur à leur emploi généralisé. Un robot sans capteur - c'est une situation encore courante dans l'industrie - a besoin d'un environnement parfaitement déterministe pour pouvoir travailler. Ainsi les objets à manipuler doivent lui être fournis avec une grande précision de positionnement. Toute situation imprévue, telle que l'alimentation de fausses pièces ou la présence d'obstacles dans le champ de travail, peut conduire à des détériorations de l'outillage, voire à des accidents. L'organisation d'une place de travail pour un robot sans capteur s'avère généralement coûteuse.

Lorsqu'on allie le sens visuel à un robot, des caractéristiques spécifiques apparaissent, qui sont plus ou moins favorables. Parmi les avantages, il faut noter la possibilité d'intervenir activement dans la scène, par exemple en montrant sous un angle adaptatif un objet à la caméra; possibilité également de contraindre mécaniquement les objets de manière à limiter l'incertitude de leur pose; enfin, des capteurs complémentaires de contact et de proximité montés sur l'organe de préhension peuvent contrôler le robot durant la phase finale des mouvements, ce que la vision fait très difficilement (fig. 4.21).

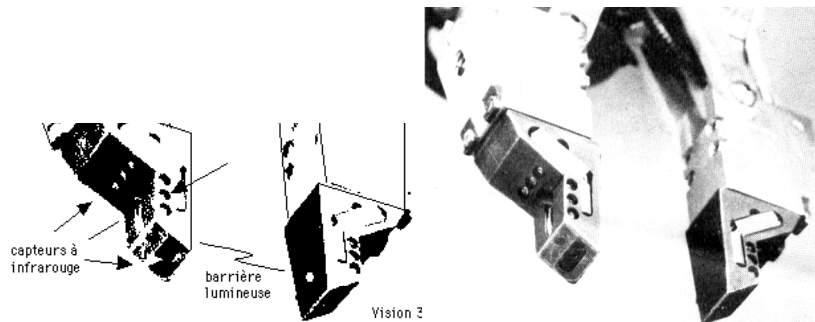


Fig. 4.21 Pince à ouverture programmable. Pour la phase terminale des mouvements, cette main possède un capteur de contact et neuf capteurs de proximité à infrarouge. Elle permet une correction fine (dx , dy , dz) au point (x, y, z) choisi par le sens visuel.

Malheureusement, des difficultés nouvelles apparaissent aussi lorsqu'on couple la vision artificielle au contrôle d'un robot. En particulier, le problème cinématique inverse est difficile à résoudre: la position et l'orientation d'un objet dans l'espace étant connues (par exemple, à l'aide du sens visuel), quelles consignes doit-on donner aux n articulations du robot pour qu'il y accède (typiquement, n vaut 6). Ce

problème d'apparence anodine n'est pas toujours soluble en pratique, que ce soit à cause de la structure du robot, de difficultés de calibration ou du temps de calcul élevé; c'est pourquoi la majorité des robots industriels actuels se programment dans l'espace des variables d'articulations et non dans un référentiel cartésien (pour la position) avec, par exemple, les angles d'Euler pour l'orientation.

La vision artificielle permet d'adapter les robots industriels à certaines variations de leur environnement. Les temps de calcul liés à la vision et au contrôle, qui sont typiquement de l'ordre de la seconde, permettent de planifier des trajectoires adaptatives. Par contre, ces temps sont prohibitifs s'il s'agit de contrôler le robot le long de la trajectoire voulue. Ainsi, un vieux rêve de roboticien n'est toujours pas réalisable: Contrôler par la vision une structure mécanique peu rigide, pour en faire un robot de haute précision.

6 MATERIEL POUR LA VISION

Le matériel idéal pour l'inspection comprend, en particulier, un capteur à haute résolution. Pour une grande vitesse de traitement, une unité câblée de prétraitement a longtemps été utile, qui soit capable de gérer des fenêtres (segments d'image), et à l'intérieur de chacune d'elles, de faire du filtrage, des rotations et des changements de résolution (interpolation et rééchantillonnage) programmables. Mais avec les progrès des ordinateurs (fréquence de travail plus élevée, parallélisme, et mémoires-cache de plus en plus systématiques) l'avantage des pré processeurs matériels n'est plus évident.

Ce paragraphe décrit trois types de solutions utilisés au laboratoire de robotique de l'EIVD.

6.1 Systèmes standard

Alors qu'autrefois, la transmission d'un signal analogique vidéo était de règle entre caméra et ordinateur, avec la plupart du temps une carte spécialisée de numérisation, aujourd'hui nous avons déjà au niveau de la caméra le passage au signal numérique. Dès lors la communication peut se faire selon des normes plus générales, adoptées également pour d'autres signaux informatiques que l'image : normes Ethernet, TCP/IP, USB, firewire, Camera Channel...

6.2 Carte de numérisation pour PC: Matrox

Au LaRA un prototype original commandé à KuKam a eu le mérite de rendre possible tôt (dans les années 80) une vision artificielle déjà efficace. Mais il s'agissait d'un prototype. Il n'est donc pas facile de transporter les logiciels développés à l'époque sur d'autres systèmes.

Actuellement, le monde des PC s'est énormément développé, et il n'est pas surprenant que l'on y trouve en particulier un grand nombre de ressources pour le traitement d'images.

La fig. ci-dessous présente le schéma-bloc d'un numériseur d'images Matrox. Le système se branche comme carte sur le bus d'un PC. Relativement peu coûteux, il s'intègre facilement dans un contexte matériel et logiciel familiers aux étudiants.

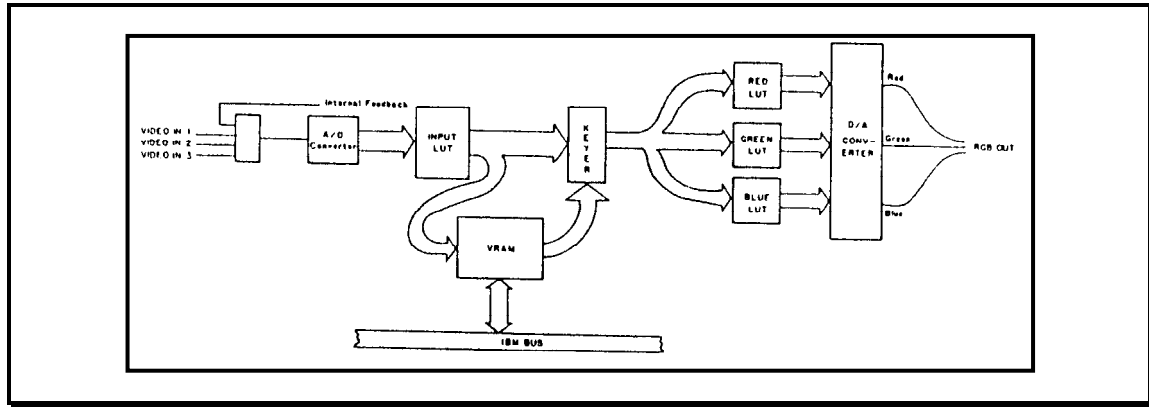


Fig. 4.24 Numériseur d'images Matrox PIP-1024 (1 million de pixels)

6.3 Traitement d'images par d'autres moyens

Au LaRA, nous avons utilisés maintenant en principe C++ Builder pour écrire nos algorithmes et faire de la vision par ordinateur. L'interfaçage se fait en principe par les canaux de communication standard, non spécifiques à la vision (Caméras Ethernet, USB, ou autre)

Mais nous disposons aussi de systèmes comprenant une carte spéciale de communication, pour divers systèmes spécialisés en termes d'environnement interactif (Matrox Meteor 2 et logiciel Inspector), d'autonomie en contexte industriel (Siemens VS 701, à processeur embarqué et liaison Profibus) ou encore de très haute vitesse (Olympus Encore ; 8000 images par seconde). Mais la tendance est aux liaisons rapides selon des standards génériques (USB, Ethernet-TCP/IP, Firewire) et les nouveaux systèmes sont de ces types. Pour les performances extrêmes en termes d'acquisition d'images (résolutions spatiales supérieures à 1Mégapixels par image, et temporelles inférieures au centième de seconde), les normes "camera link" s'est établie.

7 EXEMPLES D'APPLICATION

La partie du cours traitant de la vision se termine avec quelques exemples d'applications. Les applications retenues tirent leur valeur exemplaire de diverses qualités: vitesse de traitement (tri de chocolat), robustesse (guidage pour soudure à l'arc), prise en compte de 3 dimensions (contrôle de circuits imprimés), de leur caractère général (manutention d'objets) ou encore de leur disponibilité commerciale à large échelle (ABB).

7.1 Tri de chocolats

Nous avons parlé au §3.5.E du robot ARIA. On le voit dans la figure ci-dessous localiser des pralinés, à la cadence d'environ 90 cycles par minute. La prise de vue se fait en diascopie, avec une caméra à semi-conducteur. L'analyse se déroule sur une carte à transputer de type Autronic déjà présentée dans ce chapitre.

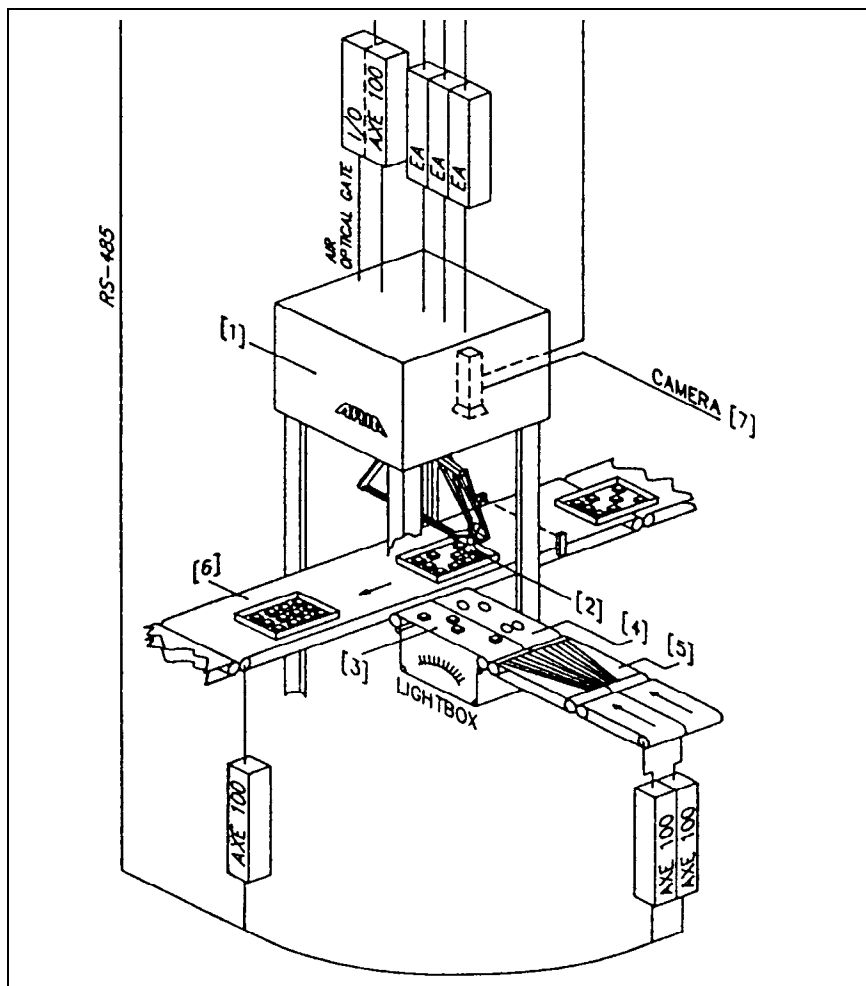


Fig. 4.28 Prise d'objets avec guidage visuel. Temps de cycle bien inférieur à la seconde.

7.2 Contrôle de circuits imprimés

Depuis une dizaine d'années déjà, chez Hitachi au Japon, les circuits imprimés s'analysent visuellement de façon à détecter les mauvaises soudures.

L'expérience a montré que l'information de forme est importante, aussi c'est une analyse tri-dimensionnelle qui est adoptée.

Par triangulation, la forme des soudures est détectée. Deux dimensions de balayage sont assurées naturellement par les caméras à semi-conducteur utilisées. En outre, dans cette application, un mouvement de translation est apporté entre les circuits inspectés et l'ensemble caméras / plan lumineux.

7.3 Guidage visuel en soudure à l'arc

Le procédé de soudure à l'arc gagne à s'automatiser, à cause des conditions de travail difficiles pour les travailleurs et à cause des coûts actuels de fabrication.

La vision est très utile dans ce procédé de fabrication, car en fonction de la forme du métal en fusion, et de l'aspect local des pièces, il est possible d'adapter la trajectoire de la torche de soudure, ainsi que sa vitesse d'avance.

Mais la soudure à l'arc pose un défi particulier à la vision, car l'éclairage varie fortement (arcs intermittents) et le milieu est agressif (fumées, jets de matière, chaleur...).

Deux exemples peuvent être cités dans ce contexte: Dans un institut de recherches de Karlsruhe, l'image est saisie durant les quelques dizaines de millisecondes où le contact existe entre torche et métal: cette situation se répète très souvent, et à ce moment là il n'y a pas d'arc. Chez General Electric, aux Etats-Unis, c'est un filtre à interférence (bande passante très étroite dans le domaine des fréquences lumineuse) qui rejette la lumière de l'arc. L'éclairage à cette fréquence (c.-à-d. à la couleur rose du filtre) est apporté par un laser.

Dans les deux applications, l'image de la scène est transporté vers la caméra par une torche optique à milliers de fibres.

En 1993, un fabricant de robot présentait à Zürich (Industrial Handling) une alternative où plutôt que d'analyser l'image, c'est les variations de courant dans l'arc en fonction de l'éloignement de la tôle qui servait de source de correction de trajectoire.

7.4 Manutention d'objets

Les opérations de manutention de pièces se déroulent schématiquement en trois phases: acquisition, transport et positionnement.

Durant l'acquisition, le type et la pose des objets à prendre sont parfois parfaitement déterministes. Cette situation est particulièrement favorable au robot et le dispense d'avoir un sens visuel. Cependant, il n'est pas toujours économique d'organiser cet environnement (coûts de stockage et de transport des pièces orientées, magasins spéciaux, équipements de mise en ordre après ébavurage, p. ex.). Parfois, les pièces ont une pose aléatoire, mais elles sont présentées une à une. C'est souvent le cas lorsque les pièces arrivent sur une bande transporteuse. Le robot peut alors en estimer la pose et les acquérir en conséquence. Mais, il y a dans l'atelier, beaucoup de pièces stockées en vrac. Traditionnellement, c'est

de la main d'œuvre ou des bols vibreurs qui les manipulent pour en alimenter des machines, des magasins ou des supports divers. Dans bien des cas, il apparaît maintenant possible de céder cette tâche à un robot doté de vision. Le sens visuel se limite alors à la détection de points de prise, laissant l'estimation de pose à plus tard, lorsque la pièce est présentée de manière plus adéquate au capteur d'images.

Durant le transport, l'objet peut bouger dans l'organe de préhension et une estimation de pose peut être à nouveau requise. A l'opposé, l'incertitude de pose peut être restreinte si des contraintes mécaniques sont introduites. Ceci arrive, par exemple, lorsqu'une pièce doit être réorientée, voire changée de main pour pouvoir être déposée au suite désiré.

Le positionnement, aussi, a parfois recours au sens visuel. Ainsi, le site où la pièce doit être déposée peut ne pas être localisé a priori. Une estimation de pose du but est alors nécessaire. Il arrive même que le but bouge et qu'ainsi le robot doivent procéder à une estimation de mouvement.

Lorsqu'on analyse une place de travail, on est souvent surpris de la diversité des contributions faites par l'opérateur. Ainsi, il y a presque toujours un certain contrôle de qualité (inspection) qui est accessoire à la tâche principale. Si l'on veut remplacer l'opérateur, il est alors judicieux de prévoir un substitut à toutes les fonctions essentielles.

7.5 ABB

ABB est un grand fabricant mondial de robots (en valeur). Il est donc intéressant de voir ce qu'ils offrent en vision pour leurs applications.

Il apparaît tout d'abord que la vision n'est pas courante sur leurs systèmes. Outre les défis généraux que la vision industrielle rencontre, il s'ajoute ici le fait que la commande de leurs systèmes laisse peu d'initiative à l'utilisateur.

Par contre, deux éléments favorables résultent d'une très bonne intégration de leurs équipements: l'utilisateur dispose, sur le boîtier portable de commande, de directives relatives à la vision. Et les problèmes de changement de coordonnées sont pris en charge pour l'essentiel par l'unité de commande standard.

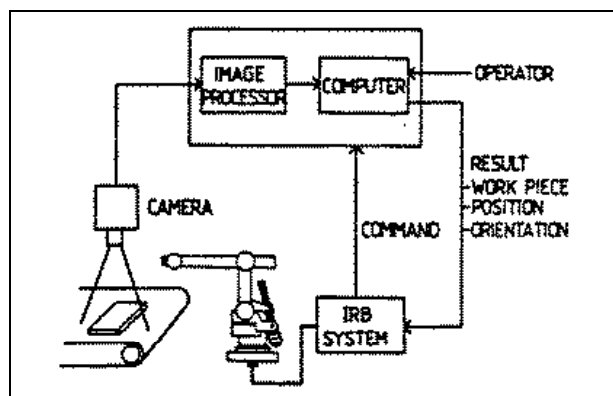
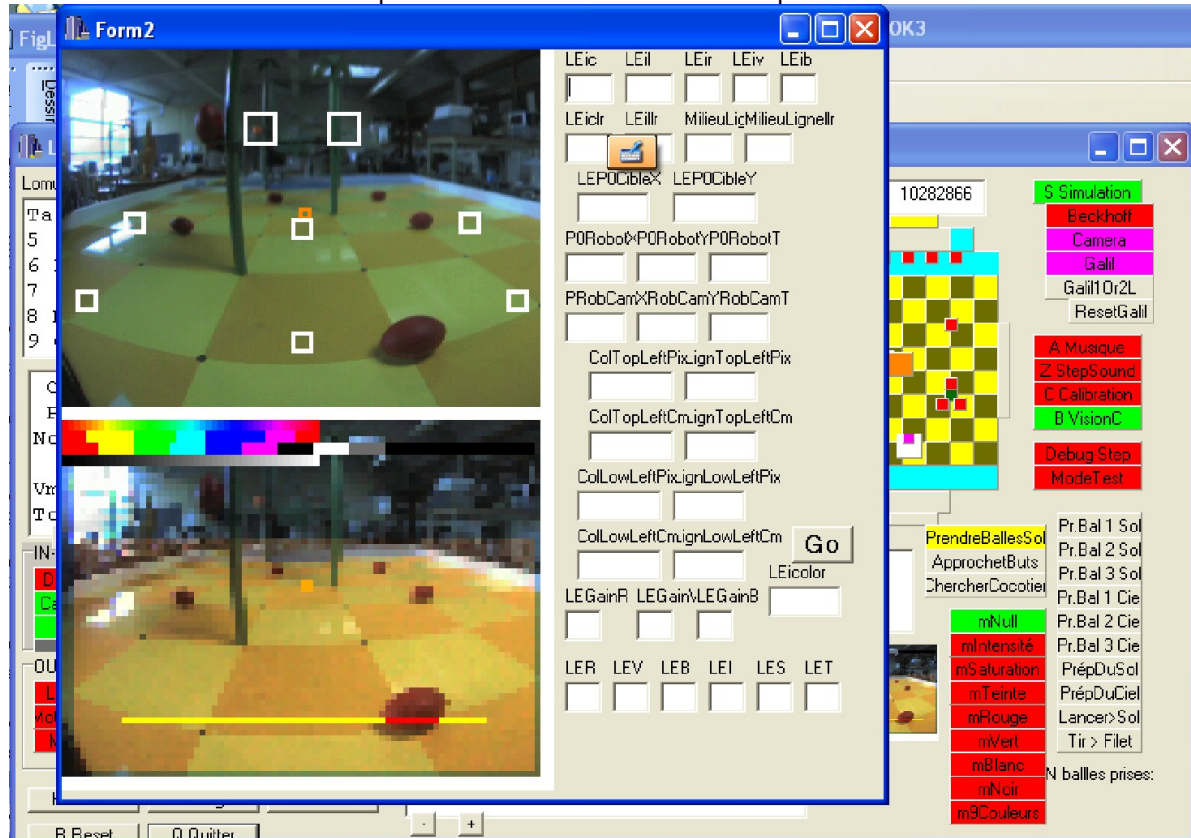


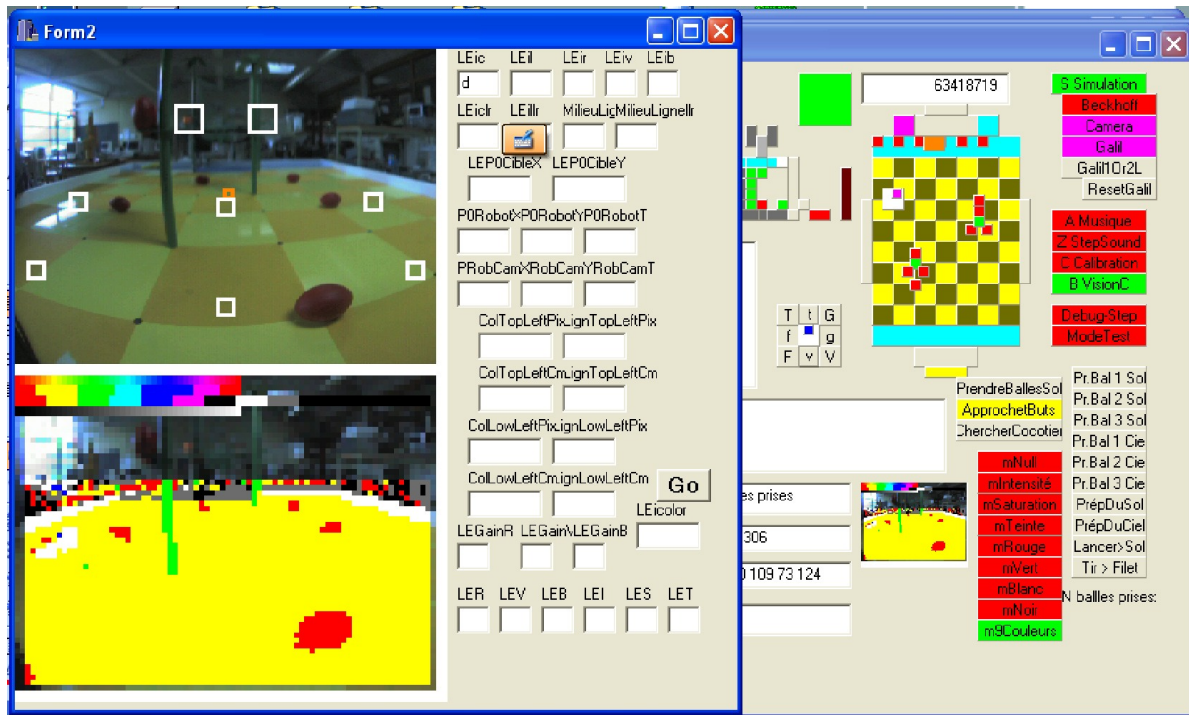
Fig. 4.29 Exemple de système robotisé comprenant en option un système de vision intégré

Un autre exemple industriel de robots dotés de vision est fourni par la firme Adept.

7.6 Robots mobiles autonomes de la HEIG-VD

Les robots mobiles autonomes de la HEIG-VD, développés d'abord pour les coupes de robotiques Eurobot, puis pour les applications en R&D pour robots de services ont toujours eu une fonction visuelle. Voici un exemple tiré de l'application « rugby » dont le robot Lomu s'occupait. Dans cette applications, de nombreuses fonctionnalités présentées dans ce cours sont implémentées.





8. CONCLUSION

La vision artificielle a commencé récemment à permettre l'automatisation rentable de certains processus industriels.

Les concepts de base ont été présentés, ainsi que certaines méthodes plus spécialisées telles par exemple celles qui interviennent pour l'acquisition de pièces en vrac, pour l'estimation de pose des objets, pour l'évaluation de distance ou pour le contrôle de qualité.

Dans bien des situations cependant, la vision artificielle n'offre pas encore de solution viable. La plus grande difficulté, c'est généralement la saisie de la scène à analyser de façon à obtenir une bonne représentation en mémoire. Un autre obstacle est posé par la quantité d'opérations que le traitement bidimensionnel des images requiert. Ceci implique souvent un matériel de grande complexité ou alors, si l'essentiel est fait par logiciel, beaucoup de temps de calcul. Enfin des limitations proviennent du manque de généralité des algorithmes existants, du défaut de modèle géométrique adéquat (lorsqu'elle guide un robot), et du peu d'expérience disponible à l'intérieur des entreprises, en vision artificielle.

REFERENCES BIBLIOGRAPHIQUES

Général

- R.C. Gonzalez and P. Wintz, "Digital Image Processing, 2nd ed., Addison-Wesley, 1987.
- D. Marr, "Vision", W.H. Freeman, New York, 1982
- M. Kunt, "Traitement numérique des signaux", vol. XX du Traité d'électricité, PPR.
- T. Pun, "Vision artificielle: apport des méthodes classiques", Intelligence artificielle et robotique, DIP et FDM, Sion, 1989.
- Christopher D. Watkins, Alberto Sadun, Stephen Marenka, "Modern image processing : warping, morphing, and classical techniques" : Boston (etc.), AP Professional, 1993, pp. 234 + 1 Diskette ISBN 0-12-737860-X
- Harley R. Myler, Arthur R. Weeks, "The pocket handbook of imaging processing algorithms in C" , Englewood Cliffs, New Jersey : PTR Prentice Hall, 1993, pp. 303, ISBN 0-13-642240-3
- V. V. Alexandrov and N. D. Gorsky, "Image representation and processing", Kluwer Academic, 1993
- Milan Sonka, Václav Hlavac and Roger Boyle, "Image processing, analysis and machine vision" London (etc.) : Chapman & Hall (Chapman and Hall computing series), 1993, pp. 555, ISBN 0-412-45570-6
- J.-D. Dessimoz, A. Beran, S. Ernst, O. Olmo and L. Venries, "Two Antinomial Methods For Automatically Solving Assembly Tasks; Case of Picking and Stacking Film Cans", Proc. 1994 Latsis Conference: From Perception to Action (Per'Ac), at EPFL, Lausanne, Switzerland, Sept. 4-9, published by IEEE Comp. Soc., New York, 1994 (winner of Robot Contest).
- J.-D. Dessimoz, P.-A. Schüpbach, A. Mercier and T. Pun, "Do I need a robot or a nonrobot automated system?", Proc. IEEE/Rob. Soc. of Japan Conf. on Intelligent Robots and Systems (IROS'92), Raleigh, North Carolina, USA, July 7-10, 1992.
- T. Pun and J.-D. Dessimoz, "Computer interpretation of complex tridimensional scenes, with application to the visual guidance of industrial robots", National Research Program NFP/PNR 23 - Artificial Intelligence and Robotics, Program Status Report 2, Bern, July 1, 1991.
- J.-D. Dessimoz, "Artificial Vision for Industrial Robots", in the book Applications in Artificial Intelligence, S.J. Andriole Ed., Petrocelli Books, Princeton, New-Jersey, 1985, pp353-369.
- J.-D. Dessimoz and P. Kammenos, "Software or Hardware for Robot Vision", in the book Artificial Vision for Robots, I. Aleksander Ed., Kogan Page Ltd, London, 1983, (ISBN 0 85038 757-41).
- J.-D. Dessimoz, P.-F. Gauthey, L. Etorre, G. Genna J. Pittet and J.-Ph. Thiran, « Rapid Prototyping for Visual Analysis Tasks », Workshop national IPLnet 2003 ; Challenges of Globalization for Swiss Manufacturings and Logistics », Ebnat-Kappel, Sept. 8-10, 2003, CD-ROM ISBN 03-033-00059-2, C.Heitz et al. Eds., Réseau national de compétences en Productique et Logistique Intégrales IPLnet, ZHW-IDP, Winterthur, mars 2004. (voir aussi sur le web. cf : www.iplnet.ch)

Acquisition d'images

http://www.infraredtraining.com/ir_primer.asp (cf. température et microbolomètres)

Extraction et traitement de contours

- L.G. Roberts, "Machine perception of three dimensional solids", Optical and Electro-optical Information Processing, Tipett, J.T. et al. Eds) MIT Press, Cambridge USA, pp159-197, 1965.
- J.-D. Dessimoz, "Sampling and Smoothing Curves in Digitized Pictures", (Tutorial), First European Signal Processing Conf., Lausanne, Sept. 16-19, 1980, pp157-167.

Estimation de distances

Thierry Oggier, « Swiss Ranger SR-3000 3D-TOF Camera », Imaging.CSEM, Zurich, Switzerland, March 2006, 26pp. (see also www.swissranger.ch)

Reconnaissance de formes

H. Niemann et al., "ERNEST: A Semantic Network System for Pattern Understanding", IEEE Trans. on Pattern Analysis and Machine Intelligence, PAMI-12, No 9, Sept. 1990, pp.883-905.

J.-G. Postaire, "De l'image à la décision", Dunod -informatique, Bordas-Paris, pp.186,1987

Texture

R.M. Haralick, "Statistical and structural approaches to textures", Proc. of IEEE, Vol.62, No5, pp 786-804, 1979

M. Unser and F. de Coulon, "Detection of defetcs by texture monitoring in automatic visual inspection, 2nd Int. Conf. on Robot Vision and Sensory Control, Stuttgart, Germany, pp.27-38, 1982

Prise de pièces en vrac par robots et robots mobiles

R. Kelley, H. Martins, J. Birk and J.-D. Dessimoz, "Three Vision Algorithms for Acquiring Workpieces from Bins" (invited paper), in Selected Papers on *Industrial Machine Vision Systems*, SPIE Milestone Series, B. G. Batchelor and P. F. Whelan Eds, SPIE Press, USA, pp. 497-514, 1994.

J.-D. Dessimoz et al., "Matched Filters for Bin Picking", IEEE Trans. on Pattern Analysis and Machine Intelligence, New York, vol 6, No 5, Sept. 84

Nicolas Uebelhart, Florian Glardon and Pierre-François Gauthey, "Lomu, an Autonomous Mobile Robot with Robust Architecture and Components", « DARH-2005 - 1st International Conference on Dextrous Autonomous Robots and Humanoids», with sponsorship Eurobot, IEEE, CLAWAR, and CTI, HESSO-HEIG (West Switzerland University of Applied Sciences), Yverdon-les-Bains, Switzerland, May 19-22, 2005. (re: [DARH2005](#)) [AbstractWithFig \(1.3Mb\)](#)

Compression/codage)

James A. Storer (ed.), "Image and text compression", by Boston (etc.) : Kluwer Academic Publishers, 1992, 354 S. : Ill. ; 25 cm (The Kluwer international series in engineering and computer science ; 176. Communications and information theory) ISBN 0-7923-9243-4 ETH-BIB P 717 171: 176

William B. Pennebaker, Joan L. Mitchell, "JPEG still image data compression standard", New York : Van Nostrand Reinhold, 1993, pp. 638, ISBN 0-442-01272-1

Edward R. Dougherty, "An introduction to morphological image processing" Bellingham, Washington USA : Spie Optical Engineering Press, 1992, pp. 161, (Tutorial texts in optical engineering ; vol. TT 9) "A Publication of SPIE - the International Society for Optical Engineering" ISBN 0-8194-0845-X

Majid Rabbani, ed., "Selected papers on image coding and compression", Bellingham, Washington USA : SPIE Optical Engineering Press, 1992 , pp. 678 (SPIE milestone series ; vol. MS 48) "A publication of SPIE - The International Society for Optical Engineering" ISBN 0-8194-0888-3 (hardbound). ISBN 0-8194-0889-1 (softbound)

INDEX

"superrésolution"	19	Analyse de scènes	58
"zoom"	13	analyse par régions	78
Aberrations chromatiques	15	analyse tridimensionnelle	86
abscisse curviligne	65	angle d'ouverture	14
acquisition d'images	4, 6	applications	3
adressage	51	approximation polygonale	65
algorithmes heuristiques	78	Arbres de décision	71
Amélioration d'image	38	axe d'inertie	61
analyse d'images	4	balayage	17, 19
analyse d'images par machines	37	Blooming	21
analyse de connectivité	66	bols vibreurs	79

bougé	8	courbure du contour	77
but de l'acquisition d'image	7	critère de Bayes	71
cadence d'acquisition	16	cyan	32
Caméras à semi-conducteurs	16	débit d'information	1
Caméras à tube	15	décomposables	40
capteur bidimensionnel	16	défocalisation de l'objectif	41
capteurs linéaires	17	déformations "élastiques"	51
caractéristique	70	Déformations "esthétiques"	51
caractéristiques	59	Détection de mouvement	56
Caractéristiques spectrales	7	détection de pic	74
Carte de numérisation pour PC	83	détection des mouvements	55
CCITT	54	diaphragme	14
cellules défectueuses	16	diascopie	8
centre de gravité	60	diffus	8
circuits intégrés	74	dilatation	43, 47
circuits intégrés spécialisés	74	distance	72
classification	2, 68	Distance	70
Clipping	49	distorsions	71
C-M-Y-K	31	Distorsions géométriques	15, 20, 51
Codage	25	distorsions géométriques de l'image	16
Codage de la couleur	22	droite des pourpres	31
codage polaire	64	Durée de vol	72
code de chaîne	63	dynamique	20
code de plage	63	éblouissement	17, 21
Cognex	51	échantillonnage	20
Compensation des .i.distorsions géométriques	51	Échantillonnage	22
Complexité	10	éclairage artificiel	7
composantes R, V, B	30	éclairage homogène	8
Composition des couleurs	26	éclairage uniforme	7
compression	52, 65	effet de ballonnement	20
connectivité	66	Effet de Moiré	20
Contrôle de circuits imprimés	85	effet de vagues	51
contrôle de qualité	87	effets de perspective	51
contrôle de robots	2	élément structurant	46
Conversion de .i.texture	48	ellipse d'inertie	61
convexité	71	entrelaçage	16, 23
convolution	40	épiscopie	8
coordonnées	61	Érosion	43, 46
Corrélation	68	espace des caractéristiques	70
corrélacion d'images	74	Espace des changements d'échelle	48
Couleur	9	espace des poses possibles	77
couleur "intrinsèque"	30	estimateurs	71
couleur normalisée	28	estimation de .i.pose	76
courbure	64, 65	Estimation de distance	72

estimation de pose	77	fréquence <i>spatiale</i>	11
État de surface	9	fréquences spatiales	12
étiquetage	66	Gain automatique	20
Exemple de numérisation d'image	25	gain fixe	20
Exemples d'application	85	géométrie d'une scène	2
Extraction de .i.caractéristiques	59	grain	14
Extraction de bords	41	Guidage visuel en soudure à l'arc	86
extrapolation	49	haute définition	23
fac-similé	52	haute résolution	78
facteur d'échelle	13	Histogramme	59
faire ressortir un objet	9	histogramme des niveaux de gris	59
Faux contours	24	Histoire	5
fax	52	homogénéité	10
fenêtres	49	hue	30
fermeture	47	identification	77
fibres optiques	7	image booléenne	60
filtrage passe-bas	15	image nette	14
filtre "analyseur"	12	infographie	2
filtre "polariseur"	12	infrarouge	11, 12
filtre à interférence	86	inspection	2, 78
Filtre médian	44	intensité lumineuse	30
filtre passe-bas	40	interférométrie	7, 76
filtre spatial	81	invariant	64
filtrer le contour	65	ITS	31
filtrer passe-bas	66	JPEG	52
Filtres	11	Kanade	74
Filtres adaptés	42, 81	kernel	40
filtres couleurs	11	labeling	66
filtres gélatines	12	Labelisation	49
filtres interférométriques	12	Laplaciens	40
Filtres linéaires	39	laser	7
Filtres non-linéaires	43	lignes de lumières	8
flou	14	limitations d'utilisation des capteurs visuels	2
<i>flux</i> d'images	52, 55	lissage	65
Focale	12	logiciel Photostyler	51
focaliser	12	loi de Lambert	9
fonction de .i.courbure	64	longueur d'arc	65
fonction de probabilité	59	longueur effective	60
Fonction de transfert	18	Look-up table	38
Fonctions de transfert	37	lumière ambiante	7
forme des soudures	86	lumière cohérente	7
forme tridimensionnelle	9	lumière monochromatique	7
formes naturelles	71	lumière naturelle	7
fréquence d'échantillonnage	23	magenta	32

manutention	86	Perception à distance	2
manutention d'objets	3	périmètre	60, 77
Marche à suivre pour acquérir une image	32	perspective	71
Marr	5	perte de contenu spectral	41
Matériel pour la vision	82	pixels	22
Matrox	83	plages	52
Méthode des Différences	69	plan focal	13, 14
Méthode des Différences de couleurs pondérées selon saturation	69	plans lumineux	8
Méthodes statistiques	68	plus proche voisin	71
Méthodes structurales	71	poids des éléments du bord	60
Méthodes topologiques	72	point de prise	80
milieu de transmission	6, 10	point par point	36
mise au net	72	points particuliers	62
mise au point	12, 15, 76	polarisation	11, 12
mode de représentation	63	polarisée	9
Modifications des pixels d'après leur voisinage	39	pose	76
Modulation	8	position d'un objet	60
moments d'inertie	61, 77	Position de la source	8
Morphologie mathématique	46	Poursuite de contours	67
Mouvement	10	poursuite des contours	63
Moyenneur	40	précision	60
MPEG	52	Principes physiques de transduction	18
néon	7	prise de pièces en vrac	79
niveaux de gris	24	Prismes anti-bougé	11
nombre de cellules	16	problème cinématique inverse	82
Nombre de lignes	22	Profondeur de champ	14
Nombre de points par ligne	23	Pseudo-couleurs	38
noyau	40	puissance de l'image	76
Objectif	12	Quantification	24, 38
observateur humain	37	radiométrie	72, 76
obturateur rapide	8	reconnaissance des formes	68
oeil humain	38	recouvrement	77
opération de filtrage	40	recouvrement spectral	20
Opérations globales	49	récuratif	81
Opérations par pixels individuels	36	Rééchantillonnage	50
Opérations sur des .i.fenêtres	49	Rééchantillonnage	23
orientation	61, 64	Réflectance	9
Ouverture	14, 47	réflexion spéculaire	8
palettes de couleurs	38	régions	49
parallaxe	51	relaxation	47
Passe-bas spatial	12	Rémanence	21
pattern recognition	68	Réorganisation du champ visuel	51
pels	22	réponse impulsionnelle	40
		représentation polaire	64

résolution	14, 72	Techniques de .i.compression	52
résolution spatiale	15, 19	teinte	30
résolution temporelle	21	teinte dominante	31
ressemblance	68	télévision	22
Roberts	40	temps d'analyse	78
Rotation	50	temps d'exposition	10
rugosité	9	temps d'intégration	17, 21
Run-length encoding	52	Temps de traitement	56
RVB	31, 32	temps de vol	72
saturation	21, 30	texel	10
scale-space	48	texture	10, 48
scène en mouvement	8	ton pastel	31
Segmentation	66	traitement bidimensionnel d'images	4
sens visuel	1	Traitement bidimensionnel	36
sensibilité	18, 21	traitement d'images	36
série de filtres	40	traitement des signaux	5
seuil de décision	59	traitement hiérarchisé	78
Seuillage	38	transformation polaire	51
Shannon	5	transformations spatiales	51
silhouette	8, 60	Transformée en cosinus	56
silicium	17	transformée en cosinus	53
source continue	8	Tri de chocolats	85
source ponctuelle	8	triangulation	72, 73
Sources de lumière	7	tubes	13
sources fluorescentes	7	valeurs propres	62
splines	66	variables d'articulations	82
Squelettisation	45	vision artificielle	2
stéréoscopie	72, 75	vision humaine	25
stéréoscopique	74	vision par ordinateur	1
stroboscope	8	vision pour robots	81
surexposition	18	vision stéréoscopique	6
surface	60	vrac	79
surface brillante	8	Wiener	5
surface circonscrite	77	YUV	31
surface mate	8	zones d'ombre	74
surveillance	3	zones d'ombres	73
synthèse	2	zoom	50
table de conversion	38		