UNIVERSITÉ LIBRE DE BRUXELLES
FACULTÉ DES SCIENCES
DÉPARTEMENT D'INFORMATIQUE

Analyse vidéo pour la détection, suivi et reconnaissance de poissons

DELLA MONICA Simon, OOMS Aurélien,SONNET Jean-Baptiste

Superviseur : Yann-Aël Le Borgne

Année académique 2012 - 2013

Table des matières

1 Introduction **2**
1.1 Motivation . . . 2
1.2 Contexte, matériaux et questions . . . 2
1.2.1 La question du *mouvement* . . . 3
1.2.2 La question de la *correspondance* . . . 4

2 État de l'art **6**
2.1 Caneva . . . 6
2.2 Segmentation . . . 6
2.2.1 Image binaire . . . 7
2.2.2 Watershed . . . 7
2.2.3 Détection de contours . . . 8
2.3 Extraction . . . 9
2.4 Analyses . . . 9
2.4.1 Matcher . . . 9
2.4.2 Distance euclidienne . . . 9
2.4.3 SURF . . . 9
2.5 Suivi . . . 9
2.6 Exigences . . . 9

3 Méthodes implémentées **10**
3.1 Conditionnal Density Propagation . . . 10
3.1.1 Notation . . . 11
3.1.2 Dynamique du système . . . 11
3.1.3 Échantillonnage . . . 12

4 Résultats expérimentaux **13**

5 Discussion **14**

6 Conclusion et perspectives **15**

Bibliographie **17**

Chapitre 1

Introduction

1.1 Motivation

L'analyse logicielle vient de plus en plus appuyer le travail de tous ceux pour qui la récolte d'informations provient de processus d'acquisition numérique, i.e des vidéos enregistrées d'observations menées par des biologistes ou statisticiens.
Le usage de procédés automatisés permet le filtrage et l'analyse massive de données jusqu'alors fastidieuse à traiter. Par exemple, l'analyse des déplacements individuels au sein d'un banc de poissons est une tache pour laquelle la puissance d'un traitement informatisé apparaît comme essentiel.
Le but de ce projet est d'ébaucher une solution logicielle d'analyse vidéo permettant d'automatiser le travail de détection et de récolte de données d'un banc de poissons dans un milieu donné. Données desquelles il sera possible d'inférer ou d'appuyer des thèses selon les besoins..

1.2 Contexte, matériaux et questions

Une vidéo est un flux, une succession d'images, dites *frames*. Le flux est dépendant de la fréquence de lecture d'affichage, soit un taux en fps (*frames per second*).
Suivre une cible au long d'une séquence d'images, c'est définir et reconnaître une partie de l'image courante comme identique à la précédente, tout en lui autorisant des modifications (taille, position, couleur,...). Rétroactivement, c'est aussi reconstruire la trajectoire d'une cible dans la séquence d'images, cette approche se montrera pertinente pour la suite.

Le *video tracking* ou suivi de cibles à partir d'une vidéo est le suivi de morceaux ciblés d'images au travers du flux duquel elles proviennent.
Si une cible apparait à l'oeil humain comme consistante alors qu'elle se meut dans son champs de vision, il en va tout autrement en *computer vision*. En effet, a tout instant la consistance conférant son unité à la cible doit être déterminée, calculée et éprouvée pour avoir une chance d'être reconnue.
Les parties traquées sont en fait chacune une quantité donnée de pixels, dans une espace

restreint. Quantité à laquelle on concédera une identité en lui définissant/reconnaissant des attributs propres (variants et invariants). La cible sera alors contenue par une *région d'intérêt*, c'est-à-dire une sous-matrice d'une matrice plus grande que serait l'image toute entière.
Le suivi proprement dit est la tache automatisée qui consiste en la localisation, d'images en images, d'un groupe de pixels généralement en mouvement.

On perçoit plusieurs difficultés qu'il conviendra de maîtriser : l'unité de la cible et sa différenciation du fond, l'identité de celle-ci, malgré des modifications intrinsèques dans le flux d'images, la caractérisation et détermination du mouvement de la cible.
Deux problématiques fortement liées peuvent être dégagées [1] : le *mouvement* et la *correspondance*.

1.2.1 La question du *mouvement*

La question du mouvement comporte deux aspects : la détermination d'une cible comme mouvante et la caractérisation de ce mouvement.

Une cible mouvante

L'approche qui semble la plus évidente pour cerner une cible en mouvement est sa différenciation avec ce qui apparait comme statique.
On supposera les objets d'intérêt et mouvant, comme appartenant à l'avant-plan (*foreground*) et le reste, statique, comme appartenant au fond (*background*). On parlera alors de *soustraction du fond*.
La soustraction du fond peut se réaliser de différentes manières, mais elle doit pouvoir résister aux changements de luminosité, aux bruits et aux différentes variations de vitesse. De façon générale, l'extraction s'obtient en soustrayant de l'image courante une image de "référence", le fond.

Un mouvement déterminé

Le mouvement est caractérisé par une position d'origine (vecteur position), une direction (orientation vers un point de destination relativement a un référentiel) et une vitesse (obtenu en dérivant les coordonnées par rapport au temps). L'enjeu est de parvenir déterminer que la position d'une cible à un instant t est la résultante d'un mouvement, de cette même cible, initié en $t-1$.

1. Video Tracking : A Concise Survey, E. Trucco, K. Plakas, IEEE Journal of Oceanic Engineering, Avril 2006

Plusieurs approches sont envisageables, la plus courante s'appuyant sur le concept de flux optique dont l'introduction est attribuée au psychologue James J. Gibson [2] .
Le flux optique (*Optical Flow*) est décrit comme le modèle sous-jacent au mouvement visible d'un objet dans son contexte et relativement à l'observateur. Ce mouvement peut être estimé à partir d'une séquence d'images comme une suite de vitesses instantanées ou de déplacements discrets d'images.

1.2.2 La question de la *correspondance*

L'objectif est de reconnaître et identifier une cible d'une frame à l'autre.

Critères d'intérêt

Pour suivre une cible, on la compare au travers du flux d'images selon certaines *unités de mesure*. Ces unités de mesure sont façonnées au regard de la problématique traitée. Elles peuvent être complexes et sont généralement hautement dépendantes des paramètres qui les composent.
Une unité de mesure est un ensemble des caractéristiques paramétrant l'objet cible, telles la position du centre de masse, l'aire, les coins, les contours ou encore l'historique des mouvements antérieurs.

La paramétrisation de la cible est capitale, car elle doit offrir des garanties sur l'identité de l'objet traqué. D'une image à l'autre, on doit pouvoir s'appuyer sur des critères robustes à l'évolution de la cible dans le flux d'image.
Il faut trouver des caractéristiques présentant des propriétés locales remarquables, c'est-à-dire des traits d'intérêts, stables ou *invariants*, on parlera de *features* de la cible.
Il existe diverses méthodes de détection de zones d'intérêts, chacune relative aux types de zones d'intérêts sur lesquels on souhaite baser l'analyse de la cible. Citons parmi les plus connu, l'algorithme de détection de coins de C. Harris et M. Stephens ou encore celui de J. Shi et C. Tomasi, mais aussi l'algorithme de détection de contours, *Canny edge detection*, présenté par l'australien J. Canny en 1986.

Fonction de mérite

Une fois la caractérisation choisie, il faut établir une méthode de comparaison débouchant sur un coefficient de qualité. Une fonction de mérite, *figure-of-merit*, qui mesure la concordance entre les données et le *modèle*, tout en considérant un choix particulier des paramètres.
En statistique fréquentielle, la fonction de mérite est généralement agencée de sorte que de petites valeurs obtenues représentent une concordance étroite. Tandis qu'une approche

2. `http://fr.wikipedia.org/wiki/Flux_optique`

bayésienne choisirait une fonction de mérite de sorte à ce que des valeurs élevées représentent une meilleure concordance[Press et al., 2007]. La fonction de mérite devra être telle qu'elle offre la meilleur façon de trouver l'extremum désiré en fonction des caractéristiques prédéfinies de la cible.
Elle devra aussi considérer que les données récoltées sont généralement bruitées. Du fait que l'objet cible se modifie tout au long du flux d'images, la reconnaissance des critères comme correspondant comprend aussi leur différenciation du contexte/bruit.
Par exemple, l'environnement où évolue la cible pourrait présenter l'une ou l'autre parcelle d'images assez ressemblante que pour passer comme identique au regard de l'unité de mesure définie. De façon générale, pour outrepasser la pollution issue d'éléments parasites valides, il sera souvent nécessaire de mettre en œuvre plusieurs approches.

Chapitre 2

État de l'art

Il existe un grand nombre d'approche pour traiter la détection et le suivi d'objet au travers de flux d'images.
Il est nécessaire de prendre connaissance de ces différentes approches et leur évolution, ainsi que des bénéfices ou contraintes qu'elles induisent. D'autre part le sujet étant prolixe, nous limiterons ici l'exposé des méthodes de détection et de suivi à celles que nous avons utilisée ou celles qui nous sont, à un moment ou l'autre, apparues importantes pour traiter de la problématique qui nous occupe.

2.1 Caneva

Même s'il n'est pas de structure commune aux algorithmes existants pour la détection et le suivi de cible en mouvement, les étapes suivantes semble prévaloir pour beaucoup de ceux-ci.

Segmentation → Extraction → Analyse → Suivi

2.2 Segmentation

A cette étape, il s'agit de dégrossir le matériau brut, de simplifier l'image en ne gardant que ce qui fait sens pour les opérations suivantes. L'objectif sous-tendant à toutes méthodes de traitement d'images étant de minimiser l'usage inutile de ressources computationnelles, la segmentation de l'image est une première approche pour ne plus focaliser que sur le signifiant.
La segmentation est une opération de partitionnement de l'image en un certain nombre de segments. Cette opération est utilisée pour dégager des zones d'intérêts de l'image. Elle assigne une étiquette à chaque pixel de sorte que tous pixels identiquement étiquetés partagent des caractéristiques visuelles données (couleur, intensité, texture). Aussi tous les segments adjacents sont sensiblement différents suivant ces caractéristiques [1].

1. http://en.wikipedia.org/wiki/Segmentation_(image_processing)

Il existe différentes méthodes de segmentation, certaines travaillent sur base de régions qu'elles accroissent, décomposent ou fusionnent, d'autres sur les contours, la classification ou le seuillage des pixels en fonction de leur intensité.

2.2.1 Image binaire

La méthode de segmentation la plus simple et la plus rapide est la création d'une image monochrome (aussi appelée *binaire*) par seuillage : À partir de l'image originale convertie en niveaux de gris, on la transforme comparativement à une valeur seuil prédéterminée en une image binaire où chaque pixel prendra une des deux valeurs possibles.
Cette transformation suivra la règle suivante :
soit une image I de taille $m * n$, un seuil T *global* et $g(x, y)$ le niveau de gris du point (x, y),

$$\forall (x, y) \in I_{\{m,n\}}, (x, y) = \begin{cases} 1 & \text{si } g(x, y) > T \\ 0 & \text{sinon} \end{cases}$$

Où les pixels appartenant à un objet de l'avant-plan sont étiquetés 1 et ceux provenant du fond sont étiquetés 0.
Cette approche grossière de seuillage peut être affinée par l'usage d'un histogramme de niveau de gris, offrant notamment la possibilité de segmenter l'image selon de multiples seuils.
Ou encore, plutôt que d'utiliser un seuil *global*, il est possible de faire dépendre T de propriétés locales du point évalué, comme par exemple de la valeur moyenne du niveau de gris de l'entourage du point considéré. On parlera d'un seuillage *adaptatif* ou *dynamique*.

L'image binaire peut, ensuite, être utilisée comme un masque permettant d'isoler des régions potentiellement intéressantes.

2.2.2 Watershed

L'algorithme de segmentation *Watershed*, traduit par "ligne de partage des eaux", se base sur une interprétation tridimensionnelle de l'image, où un point est caractérisé par ses deux composantes spatiales et son niveau de gris.
De cette image, perçue comme un relief topographique, sera calculée la ligne de partage des eaux pour délimiter le bassin-versant, c'est-à-dire l'aire à l'intérieur de laquelle convergerait de l'eau hypothétiquement tombée.
Trois types de points sont définis par cette interprétation :
- ceux appartenant à un minimum local.
- ceux à partir desquels une goutte d'eau s'écoulerait inévitablement vers une minimum local précis. L'ensemble des points relatés à un même minimum constitueront un bassin-versant de ce minimum.

– ceux à partir desquels toute goutte d'eau ruissellerait équitablement vers l'un ou l'autre minimum. L'ensemble des ces points forment topologiquement une crête, ils constituent la ligne de partage des eaux.

Il existe plusieurs d'implémenter cet algorithme, parmi les plus communes :
– selon la distance topographique d'un point au minimum le plus proche, à partir de chaque pixel de l'image, on suit le gradient jusqu'à atteindre un minimum, à l'image d'un ruissellement.
– par inondation, où est simulé une montée progressive du niveau d'eau à partir des minima du relief.

La segmentation par ligne de partage des eaux donne de bons résultats dans l'extraction d'objet presque uniforme, mais conduit souvent à une sur-segmentation dû aux bruits et irrégularités locales. Une façon de pallier à ce désavantage est d'utiliser des marqueurs. Les marqueurs sont définis comme des composantes connexes appartenant soit à un objet d'avant-plan, soit au fond. La sélection des marqueurs à garder pourra être effectuée par simple estimation du niveau de gris et de la connectivité ou par une

2.2.3 Détection de contours

Intuitivement, un contour est défini comme une suite de pixels contigus reflétant la frontière entre deux régions. La détection des contours permet alors de découper ou fusionner l'image en sous-régions.
En pratique, les principaux algorithmes de détection de contours, *edge detection*, se basent sur l'étude des dérivées de la fonction d'intensité de l'image : le gradient, les extremums locaux et le passage par zéro du Laplacien. Le contour est obtenu par détection d'une discontinuité (changement abrupte d'intensité) et des similarités (selon des critères prédéfinis).

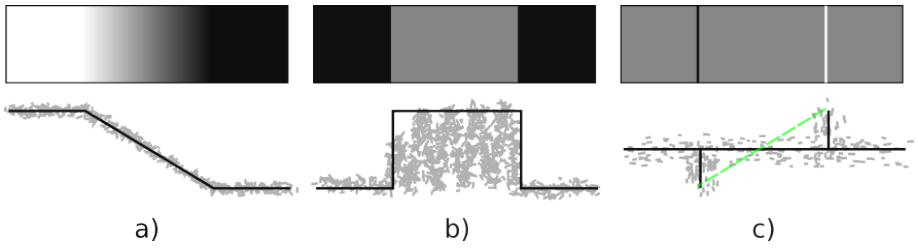

FIGURE 2.1 – Détection de contours

Dû notamment aux différentes méthodes d'acquisition, la frontière entre deux régions n'est pas toujours très contrastée ni exempte d'aucun bruit. De fait, elle apparaîtra généralement floutée et bruitée.
À partir d'une image en niveaux de gris, la frontière d'une région à une autre, représenté en *a* dans la figure 2.1, peut être *idéalement* définie par une fonction rampe dont la longueur sera caractérisé comme le niveau de floutage de la frontière.

Par l'étude de la dérivée première calculée en utilisant le gradient d'un point considéré (2.1 *b*), on peut déterminer si comparativement à un voisinage, on se trouve potentiellement sur un point du contour.
De même, en évaluant la dérivée seconde par application du Laplacien (2.1 *c*), on peut :
- en étudiant son signe, caractériser le point du contour comme appartenant à l'un ou l'autre coté de la frontière.
- déterminer le milieu exacte de la frontière floutée. En calculant la droite imaginaire (2.1 *c*, ligne pointillée) joignant les extremums de la dérivée seconde, on obtient au passage à zéro le point médian.

Le filtre Canny est un des algorithmes les plus utilisés pour la détection des contours, il se base sur l'intensité et la direction du gradient.

2.3 Extraction

2.4 Analyses

2.4.1 Matcher

2.4.2 Distance euclidienne

2.4.3 SURF

2.5 Suivi

2.6 Exigences

Suite à ce qui a été énoncé, il est intéressant de dégager quelques exigences auxquels se devrait de répondre un système de tracking robuste.
- *Faux positifs, faux négatifs et résistance à la pollution d'éléments parasites*, il convient de ne suivre que ce qui doit l'être.
- *Fiabilité quand à une possible occlusion*, il est fort probable qu'à un moment ou l'autre, la cible sera occulté par un autre élément et réapparaîtra ensuite. Le tracking doit alors rester consistant.
- *Souplesse du tracking*, celui-ci doit pouvoir suivre des éléments aux vitesses variables.
- *Stabilité*, malgré tout, le suivi de la cible doit perdurer.

Chapitre 3

Méthodes implémentées

3.1 Conditionnal Density Propagation

Afin de suivre un objet en mouvement au travers d'un flux vidéo, on doit être en mesure de déterminer à chaque *frame* la position de la cible. Cette détermination est rarement exacte car elle s'appuie sur de nombreuses mesures, pour la plus part instables. La déformation de la cible, son occlusion, des changements de luminosité, etc, sont autant de mesures dont la propension à varier aléatoirement contribue à la génération de bruit dans la détermination de la cible. On souhaiterait approcher l'hypothétique détermination réelle qui aurait été obtenue aux moyens de mesures idéales. Tout au plus, l'approche attendue devrait être en mesure de mettre en exergue le tout ou une partie de la cible comme n'appartenant pas au bruit.
Le processus de détermination se base sur un mouvement cyclique, partant d'un modèle donné dont la paramétrisation est issue d'observations antérieurs, on consolide ce modèle en évaluant sa pertinence à l'état présent, ce qui conduit à une prédiction sur l'état probable à l'étape suivant. On distingue alors deux phases, la *prédiction* et l'*observation*. La *prédiction* est basée sur modèle affiné par les informations passées. L'*observation*, ou phase de mesure, est la récolte d'informations sur l'état courant du système en vue de corriger la prédiction basée sur les mesures précédentes.

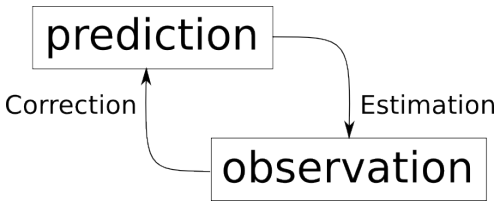

FIGURE 3.1 – Cycle de prédiction - observation

reconstruire la trajectoire d'attributs dans une séquence d'images
En probabilité, un processus stochastique vérifie la propriété de Markov si et seulement si la distribution conditionnelle de probabilité des états futurs, étant donné les états passés

et l'état présent, ne dépend en fait que de l'état présent et non pas des états passés (absence de « mémoire »). $P(\mathbf{x}_t|\mathcal{X}_{0:t}) = P(\mathbf{x}_t|\mathbf{x}_{t-1})$

3.1.1 Notation

L'*état* de l'objet $\mathbf{x}$ au temps t est noté $\mathbf{x}_t$ et son historique est l'ensemble $\mathcal{X}_{0:t} = \{\mathbf{x}_0, ..., \mathbf{x}_t\}$
De même, l'ensemble des *features* (traits *invariants*, aussi dit "les observations") est $\mathbf{z}_t$ et son historique $\mathcal{Z}_{0:t} = \{\mathbf{z}_0, ..., \mathbf{z}_t\}$
La dynamique stochastique du système (équation de transition) est entièrement donnée par $P(\mathbf{x}_t|\mathbf{x}_{t-1})$ (processus Markovien)

invariants caractéristiques locales de luminance, (photométrique) ou géométrique

3.1.2 Dynamique du système

Dynamique Stochastique L'état $\mathbf{x}$ est multidimensionnel et sa densité est plutot complexe. Pour construire un modèle dynamique, une connaissance *a priori* du mouvement est nécessaire.

Observation Les observations $\mathbf{z}_t$ sont indépendantes entre-elles et vis-à-vis du processus dynamique :

$$P(\mathcal{Z}_{t-1}, \mathbf{x}_t|\mathcal{X}_{t-1}) = P(\mathbf{x}_t|\mathcal{X}_{t-1}) \prod_{t=1}^{t-1} P(\mathbf{z}_t|\mathbf{x}_t)$$

ce qui se réduit, considérant la condition mutuelle d'indépendance des observations, en

$$P(\mathcal{Z}_t|\mathcal{X}_t) = \prod_{i=1}^{t} P(\mathbf{z}_i|\mathbf{x}_i)$$

Le processus d'observation est alors défini en spécifiant la densité conditionnelle $P(\mathbf{z}_t|\mathbf{x}_t)$ pour chaque instant t

Propagation À partir des observations, la densité conditionnelle de l'état au moment t est $P_t(\mathbf{x}_t) \equiv P(\mathbf{x}_t|\mathcal{Z}_t)$. Elle représente toute l'information de l'état pouvant être déduite de l'entièreté du flux de données.
Selon le théorème de Bayes, on déduit la règle de propagation de la densité de l'état dans le temps comme

$$P(\mathbf{x}_t|\mathcal{Z}_t) = k_t P(\mathbf{z}_t|\mathbf{x}_t) P(\mathbf{x}_t|\mathcal{Z}_{t-1})$$

où k représente une constante de normalisation ne dépendant pas de $\mathbf{x}_t$.
La densité *a priori* $P(\mathbf{x}_t|\mathcal{Z}_{t-1})$ est une prédiction issue de la densité *a posteriori* $P(\mathbf{x}_{t-1}|\mathcal{Z}_{t-1})$ provenant de l'étape précédente et à laquelle a été surimposé un pas de temps du modèle dynamique. Pour atteindre cette densité *a priori* tout en évitant un coût computationnel conséquent, celle-ci est approchée de façon récursive.

3.1.3 Échantillonnage

Algorithme d'échantillonnage

Il s'agit de retrouver un objet de paramétrisation $\mathbf{x}$ à partir d'une densité *a priori* $P(\mathbf{x})$ en utilisant les données observées $\mathbf{z}$ d'une seule image. La densité *a posteriori* obtenue par l'application de Bayes est calculée récursivement

Séquence d'images temporelles

Contrainte de flot optique Méthode différentielle construite à partir d'une formulation différentielle d'un critère de corrélation.(ex Shi-Tomasi-kanade)

méthode de corrélation Méthode basée sur des critères de corrélation (= fonction de similarité), estimer les déplacements sensible au transformations géométriques (changement d'échelle, rotation, distorsion perspective) et photométrique de l'image

Chapitre 4

Résultats expérimentaux

Chapitre 5

Discussion

Chapitre 6

Conclusion et perspectives

Bibliographie

[Canny, 1986] Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (8) :679–714.

[E. and K., 2006] E., T. and K., P. (2006). Video tracking : a concise survey. *Oceanic Engineering, IEEE Journal*, 31(2) :520–529.

[Faro et al.,] Faro, A., Giordano, D., Palazzo, S., and Spampinato, C. Fish detection and tracking. *IST – 257024 – Fish4Knowledge*.

[Fontaine et al.,] Fontaine, E., Barr, A., and Burdick, J. Tracking of multiple worms and fish for biological studies. *http ://www.cvl.iis.u-tokyo.ac.jp/mva/proceedings/2007CD/papers/11-02.pdf*.

[Gyaourova et al., 2003] Gyaourova, A., Kamath, C., and Cheung, S.-C. (2003). Block matching for object tracking. *Lawrence LiverMore National Laboratory*.

[Jain et al., 1996] Jain, A., IEEE, F., Zhong, Y., and Lakshmanan, S. (1996). Object matching using deformable templates. *Oceanic Engineering, IEEE Journal*.

[Kim,] Kim, Y. M. Object tracking in a video sequence, cs 229 final project report. *CS 229, Stanford University*.

[MAHEO et al.,] MAHEO, A.-C., Colas, R. M., and de vision, L. Méthodes de suivi d'un objet en mouvement sur une vidéo. *Département d'ingénierie et des sciences informatiques, Institut Supérieur de l'Électronique et du Numérique*.

[Moeslund, 2012] Moeslund, T. (2012). Introduction to video and image processing. *Undergraduate Topics in Computer Science*.

[Press et al., 2007] Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (2007). *Numerical Recipes the art of scientific computing*. Number 3. Cambridge University Press.

[Rova et al.,] Rova, A., Mori, G., and Dill, L. One fish, two fish, butterfish, trumpeter : Recognizing fish in underwater video. *http ://www.cvl.iis.u-tokyo.ac.jp/mva/proceedings/2007CD/papers/11-02.pdf*.

[Sellent et al.,] Sellent, A., Eisemann, M., and Magnor, M. Two algorithms for motion estimation from alternate exposure images. *Institut fur Computergraphik, TU Braunschweig, Germany*.

[Shi and Tomasi, 1994] Shi, J. and Tomasi, C. (1994). Good features to track. *9th IEEE Conference on Computer Vision and Pattern Recognition*.

[Spampinato et al., 2008] Spampinato, C., Chen-Burger, Y., Nadarajan, G., and Fisher, R. (2008). Detecting, tracking and counting fish in low quality unconstrained underwater videos. *VISAPP (2)*, pages 514–519.

[Suzuki and Abe, 1985] Suzuki, S. and Abe, K. (1985). Topological structural analysis of digitized binary images by border following. *Computer Vision, Graphics and Image Processing*.

[Wang et al.,] Wang, Y., Doherty, J., and Dyck, R. V. Moving object tracking in video. *Department of Electrical Engineering, The Pennsylvania State University, National Institute of Standards and Technology*.