|  |
| --- |
| ESGi |
| Comment diriger un module par reconstruction 3D de son environnement ? |
| Mémoire 5éme Années |
|  |
| **Augustin GARDETTE – Pascal REMINY** |
| **9/30/2014** |

|  |
| --- |
| Dans ce document nous allons analyser les différentes techniques utilisées pour la reconstruction de Scène ou d’objets 3D. |

Contents

[En bref (Introduction) 2](#_Toc403516332)

[Récupération des Données 3](#_Toc403516333)

[Single Still Image 3](#_Toc403516334)

[Markov Random Field (MRF) 3](#_Toc403516335)

[Bayesian Network 4](#_Toc403516336)

[MRF vs Bayesian 4](#_Toc403516337)

[Méthode de segmentation d’image 4](#_Toc403516338)

[Stéréographie / Stéréovision 4](#_Toc403516339)

[Camera RGB-D 4](#_Toc403516340)

[Kinect 5](#_Toc403516341)

[Traitement des Données 9](#_Toc403516342)

[Single Still Image 9](#_Toc403516343)

[Stéréographie 9](#_Toc403516344)

[Camera RGB-D 9](#_Toc403516345)

[Implémentation (Demo) 9](#_Toc403516346)

# En bref (Introduction)

Dans ce mémoire nous allons analyser les différentes technologies qui sont disponible à l’heure actuelle pour la reconstruction de scène 3D à partir de donnée brute venant d’un appareil de capture. Nous axerons nos recherches sur trois types de reconstructions, la reconstruction à partir d’une **simple image**, puis la **stéréographie** appliquée à la reconstruction d’une scène 3D, et pour finir la reconstruction 3D à l’aide de **Camera RGB-D**, type Kinect.

Chacune des parties sera traité suivant le même procédé, nous axerons dans un premier temps notre analyse sur une rapide description des techniques de collecte de données, puis sur le traitement de ces données que l’on peut récupérer grâce aux appareils présentés plus haut, puis nous vous présenterons notre propre solutions de traitement de données que nous comparerons avec les solutions existante ( ?) .

# Analyse de l’existant

Le projet Amazon Air Prime.

# Récupération des Données

## Single Still Image

Dans cette partie nous nous concentrerons sur la reconstruction d’une scène 3D à partir d’une seul image, en effet la reconstruction de scène est une problématique récurrente en vision par ordinateur, mais la plupart des travaux se focalisent sur la reconstruction de scène à partir d’un vision binoculaire de la scène, modéliser par deux image.

Il existe d’autres techniques, qui utilisent deux ou plusieurs images ou encore le dé-focus, que nous vous présenterons au fils de ce mémoire.

La plupart des techniques d’estimations de la profondeur utilisent le **Markov Random Field** ou le **Bayesian Network**. Nous allons donc dans un premier temps vous expliquer brièvement de quoi il s’agit.

### Markov Random Field (MRF)

Le MRF, ou champ aléatoire de Markov, est un ensemble de variables aléatoires caractérisées par un graphe non-orienté. Chaque arrêt représente une dépendance entre deux variables aléatoires.

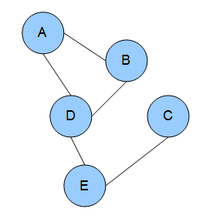


Figure  : Exemple de Markov Random FIeld

A dépends de B et D. B dépends de A et D. D dépends de A, B, et E. E dépends de D et C. C dépends de E.

Pour d’un ensemble de variable aléatoire soit un MRF il faut que ces variables aléatoires répondent toute à une propriété de Markov.

Un ensemble de variables aléatoire possède une propriété de Markov si et seulement la distribution conditionnelle de probabilité des états futurs, sachant les états passés et présent, ne dépend en fait que de l'état présent et non pas des états passés (absence de « mémoire »).

Le MRF est souvent utilisé dans l’intelligence artificielle lié à la vision par ordinateur ou au traitement d’image.

### Bayesian Network

### MRF vs Bayesian

Le MRF et le Bayesian sont similaire dans leur représentation des dépendances.

Le réseau de Markov est représenté par un graphe non-orienté alors que le Bayesian est, quant à lui représenté par un graphe non orienté.

Autre différence le MRF peut être cyclique contrairement au Bayesian qui lui est acyclique.

### Méthode de segmentation d’image

## Stéréographie / Stéréovision

## Camera RGB-D

Les caméras RGB-D ou encore caméra à temps de vol sont des caméras capable de calculer les profondeurs des objets ce trouvant devant elle. Le principe est simple une pulse de lumière est envoyée en calculant le temps du trajet en entre le premier objet rencontré et la camera (le temps de vol), on connait la distance entre la camera et l'objet. Chaque pixel de la camera effectue indépendamment cette opération et on obtient une image complète en 3D.

****



La Mesa imaging swissRanger 4000

Elle a une portée de 5-8 mètres, résolution 176 x 144 pixels sur 43,6 ° x 34,6 ° champ de vision. Elle fonctionne jusqu'à 54 images par seconde, et coûte environ $ 4300.

Dans cette partie nous allons vous expliquer comment ces hardwares fonctionnent en décrivant la fonction de l'un de ces appareils le plus connu et répandu le Kinect.

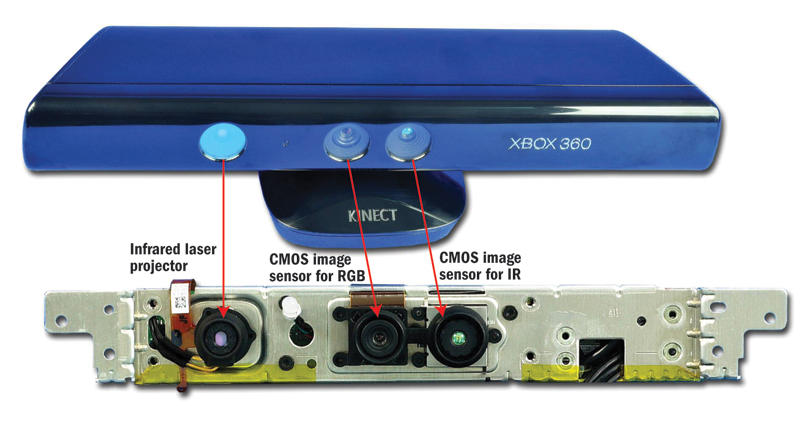
### Kinect

Dans la famille des Caméras RGB-D, nous trouvons le Kinect de Microsoft un hardware qui combine et embarque un grand nombre de technologies différentes. Il est l'un des plus connu et surtout le plus répandu sur le marché du fait de son utilisation dans le monde du jeu vidéo. Pour comprendre comment fonctionne une caméra RGB-D nous allons rentrer dans le détail des composants du kinect.

Composer d'une série de multi-microphone, une caméra RGB, un émetteur infrarouge, et enfin d'un "3D depth sensor" le kinect permet la reconnaissance vocal et faciale d'un utilisateur, ainsi que de l'ensemble d'une pièce en 3D. Il possède aussi un petit moteur à sa base pour effectuer des petits mouvements vers le haut ou le bas en fonction de la position de l'utilisateur dans la pièce.

#### La camera RGB du Kinect

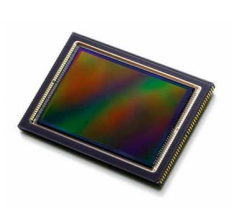
La caméra RGB est une caméra classique avec un capteur photographique de type CMOS. Elle permet une prise d'image de 32bits de couleur à une fréquence de 30Hz avec une résolution de 640x480 pixels(VGA).



**Capteur photographique** : C'est un composant électronique sensible à la lumière qui va convertir un rayonnement (Ultra-Violet, Lumière visible ou Infra Rouge) en un signal analogique (une nouvelle fois via un signal électrique).

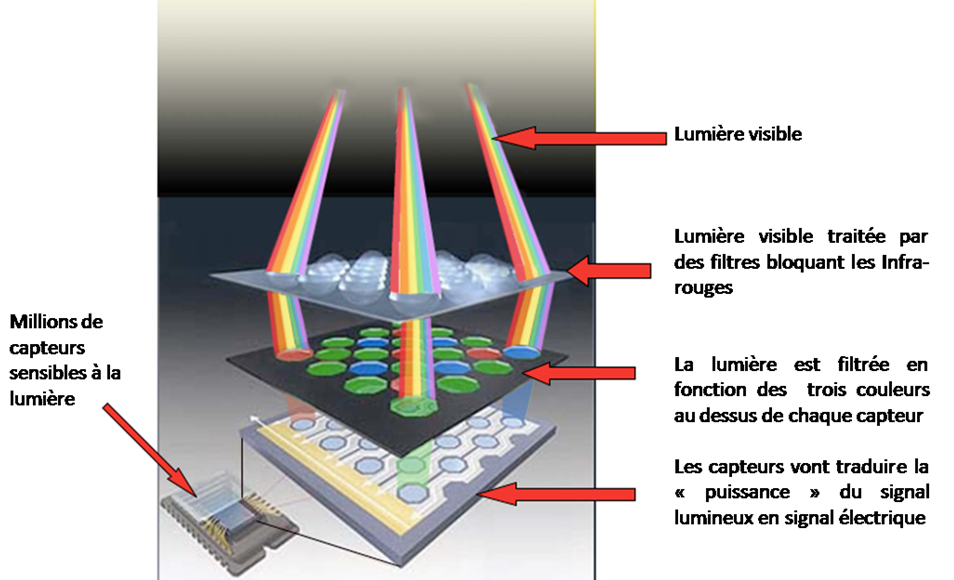
Pour convertir le signal analogique reçu en image numérique, le capteur photographique va filtrer la lumière selon les 3 couleurs primaires qui sont le Rouge, le Vert et le Bleu appeler aussi le système RVB (ou RGB en anglais) et sortir trois signaux numériques correspondant à chacune d'elles. Actuellement, deux grandes familles de capteurs sont disponibles : les CCD et les CMOS.

Les capteurs dans le kinect est de type CMOS.



**Le CMOS** : CMOS (Complementary Metal Oxide Semiconductor) est représenté sous la forme d'un petit « écran ». La partie sensible à la lumière acquiert une certaine charge électrique en fonction de la quantité de lumière reçue.

Le principe de ce capteur est simple il va tout d'abord capter la lumière ambiante, filtrer en bloquant les ultra-violet puis filtrer de nouveau la lumière avec des filtres RGB et pour finir des capteurs électrique vont  émettre un signal électrique qui se réfère à une de ces 3 couleurs afin que celle-ci soit convertie numériquement. Comme explique dans le schéma ci-dessous:

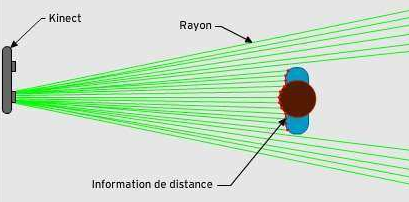


#### II- 3D depth sensor

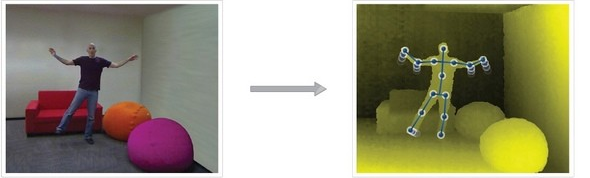
Le 3D depth sensor est la pièce maitresse du kinect. En effet c'est la partie qui va permettre au kinect de capter la profondeur et donc de création de la 3D. Pour cela kinect utilise son émetteur de lumière infrarouge qui va envoyer dans la pièce ou il se situe, une multitude de point infrarouge.



Le capteur CMOS infrarouge du kinect permet grâce à l'image qu'il obtient de calculer les distances.



De la même façon que le capteur photographique le capteur infrarouge va filtrer tout ce qui est autour de lui et cette fois ne conserver que la lumière infrarouge. Pour calculer la distance il lui suffit de regarder l'intensité lumineuse, plus un objet est prés plus l'intensité du rayonnement sera faible et donc réciproquement plus cette même intensité sera forte plus proche sera l'objet.



**Projet Tango**

Le projet Tango est un projet de recherche qui est encore en cours de développement aujourd'hui. L'idée de ce projet et simple. En effet, et si un appareil mobile pouvait avoir la vision de notre monde comme celle d'un être humain.

C'est-à-dire une compréhension à l'échelle humaine de l'espace et du mouvement comme les hommes. Cela peut être possible en intégrant les mouvements 3D complet de l'appareil, tout en créant une carte de l'environnement grâce au capteur de profondeur.

Ce projet à ouvert de nombreuse perspective sur l'avenir des Smartphones et de l'utilisation des reconstructions des scènes 3D dans notre quotidien.

Pour conclure, voici la liste des avantages et des inconvénients des caméras RGB-D:

**Avantage**

* Récupération d'une scène 3D en temps réel

Les caméras RGB-D permettent d’acquérir jusqu’à 50 images 3D/seconde des objets mesurés.

* Stabilité mécanique

Acquisition de la totalité de l’image sans procéder de balayage.

* Taille compacte

Les cameras peuvent être compactes.

**Inconvénient**

* Interférence avec des sources lumineuses

les sources de lumière ambiante peuvent interférer avec la mesure de la distance (la lumière du soleil).

# Traitement des Données

## Single Still Image

## Stéréographie

## Camera RGB-D

# Implémentation (Demo)