# 9

# The Integrated Active Stereoscopic Vision Theory, Integration and Application

Kolar Anthony[1], Romain Olivier[1], Graba Tarik[2],
Ea Thomas[3] and Granado Bertrand[4]
[1]*SYEL - University of Pierre et Marie Currie – Paris VI*
[2]*ENST - Paris*
[3]*ISEP – Paris*
[4] *ETIS - CNRS - ENSEA - Univ Cergy Pontoise*
*France*

## 1. Introduction

Monolithic integration of stereovision can make a contribution in today's solutions where main limitations are size, power consumption and deployment. Some emergent applications are not equipped with 3D capabilities although the demand is increasing.
One of these applications is 3D-endoscopy which claims good accuracy, small volume, and high autonomy in terms of power consumption and operation. Integrated stereovision used in the medical world and especially for the diagnosis of pathologies of the digestive tract gives an additional advantage for the gastro-enterologists in the observation of malignant tumors. In 3D endoscopy, integrated stereovision satisfies four essential needs:
Firstly, precision of diagnosis is required, more exactly the accuracy in the evaluation of the carcinoma size. Without a real 3D representation, estimations are done roughly due to the lack of information (depth parameter) and are defined by the experience of the expert rather than by the system itself. At the present time only an invasive exploration allows an exact location and measurement of tumor size.
Secondly, integrated stereovision meets a growing public need. Improvement in healthcare provides to the population some well-being and people live longer. The diagnosis of early forms of cancer of the digestive tract takes part in the well-being of aged people: the sooner a sick person is treated, the earlier he can recover. In the case of endoscopy, the French Company of Digestive Endoscopy (SFED) stresses in 2006 that the use of endoscopy is increasing: 2.7 million examinations were carried out, 70000 new digestive cancers were discovered and almost 900000 polyps colics were removed. Various types of endoscopy (E.O.G.D, Coloscopy, enteroscopy…) give only planar vision of the digestive tract. The contribution of the 3rd dimension would be an important advantage in the diagnosis of pathology.
Thirdly, integrated stereovision also meets a need for freedom: freedom of movement, freedom to be at home and not in a hospital. The patient swallows the capsule and he can go back to his activities. Optimization of the dimension of the capsule, autonomous operation and long battery life of the sensor are essential constraints.



Source: Stereo Vision, Book edited by: Dr. Asim Bhatti,
ISBN 978-953-7619-22-0, pp. 372, November 2008, I-Tech, Vienna, Austria

Finally, integrated stereovision meets economic needs: the number of examinations increases with the growing of old population. To optimize the cost, an early endoscopy would make it possible to improve the forecast and better cure the patient at the beginning of his disease, which reduces the costs of treatment. The sensor has to be low cost to be autonomous.
Another emergent field in which the integrated stereovision can contribute is cartography of dangerous or less accessible areas such as disaster areas, caves, surface of an asteroid or battle fields. This cartography is an aid to navigation of rescuers, robots, drones… Traditional image sensors do not allow us to know the dimension of a scene or to know the distance to the viewed objects. The interest of an integrated stereovision is that it recovers extra information characterizing the scene and its objects, in a 3D world.
For example, drones are useful in reconnaissance operations in confined surroundings, like industrial buildings, caves or houses. In these places, the drones are equipped with stereovision and can carry out photos of its surrounding area and then set up a 3D map, but it is helpful to use the furnished information to help itself for stabilization…
Another way of scoping out unknown areas quickly is to scatter a great number of stereovision sensors and thus to create a wireless network of sensors. These sensors can be easily disseminated on the location of study and supply a maximum of information for mapping danger zones or less accessible fields, like the zones of natural disasters or planets and asteroids.
The ideal 3D integrated sensor has to show a great autonomy, a good accuracy of rebuilding, good performance, a reduced size and low cost. Now, let us look the state of the art in the following chapter.

## 2. State Of Art: the integrated vision and 3D reconstruction

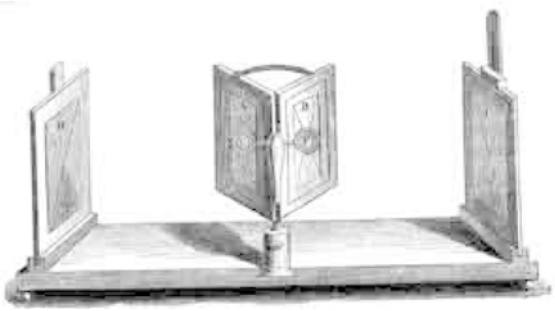


Fig. 2.1. Wheatstone's stereoscope in 1838

With the time, we had more and more needs to be able to realize a 3D representation of a scene. The advanced in the electronic vision and the image processing domain have allowed the discovery of many important solutions. The first main work in the stereovision was managed by Marr [Marr & Poggio, 1979] in the end of the year 70. These research were fixed on the calculator analyse of the human brain and particularly for the vision.
So today there are many method of rebuilding a scene in three dimensions [Song & Wang, 2006][Szeliski, 2000][Vézien, 1995][Chen & Li, 2003]. The figure 2.2 shows a ranking of the most commonly encountered methods. As we can see, there are two main groups: the reconstruction by triangulation and the reconstruction by temporal delay.

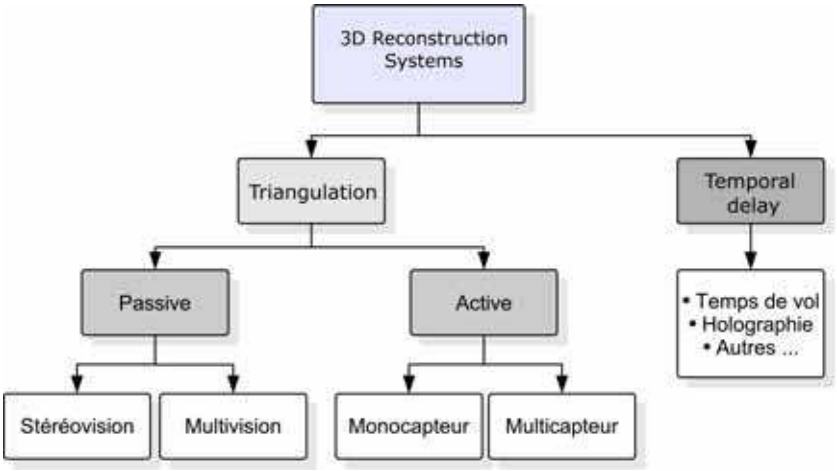



Fig. 2.2. 3D reconstruction methods classification

Since several years, the dynamics of research in this domain is such that many and robust methods to return the relief have been developed with success. For examples, the passive [Belhumeur, 1993][Cochran & Medioni, 1992] and active [Hyun & Gerhardt, 1994][Valkenburg & McIvor, 1998] stereoscopic methods or methods based on time of flight of a light pulse (monochromatic waves)[Gokturk et al., 2004][Gruss et al., 1991][Lange & Seitz, 2001]. Nevertheless, these methods are costly in material and/or temporal resources thus limiting their application in emergent domain such sensor networks for distributed 3D cartography, that could be used for the characterization of an asteroid [Stooke, 1991] and in vivo 3D endoscopy where an important depth and shape information can be added to 2D vision systems as the PillCam by GivenImaging [GivenImaging, 2005]. In fact, the continuous evolutions in the microelectronic domain allow an augmentation of the integration density. This density is such that it's possible to design integrated systems which realize more complex function on a silicon surface in constant decline. These Systems on Chip (SoC) perform the functions of a complete system. In parallel, the Vision Systems on Chip, or electronic retina [Pinna, 2003], were develop. The VSoC regroup all the needed functions for the vision. The processing capabilities and architectures that can be attached to such systems have been studied by A.Moini [Moini, 1999]. Unfortunately, sometime it's impossible have many functions on the same chip which is implement with different technology in order to optimize their performances. For this case it exist another solutions to arrive to the same level of integration. The Multi chip Component (McM) or the System in Package (SiP) [Song et al., 2003] allow to regroup on the same substrate or package many functions which is realize in different technology. A such approach allows the cohabitation of digital processing units in CMOS with telecommunication or vision unit (SiGe or AsGa).
An other form of integration was presented as the ultimate level of integration by Tummala [Tummala, 2005/2006] in order to join the Law of Moore. It is the System on Package, SoP. This approach aims to bring together the various components of technology to smaller dimensions that integrate new functions of communication while being as effective as possible (fig.2.3).

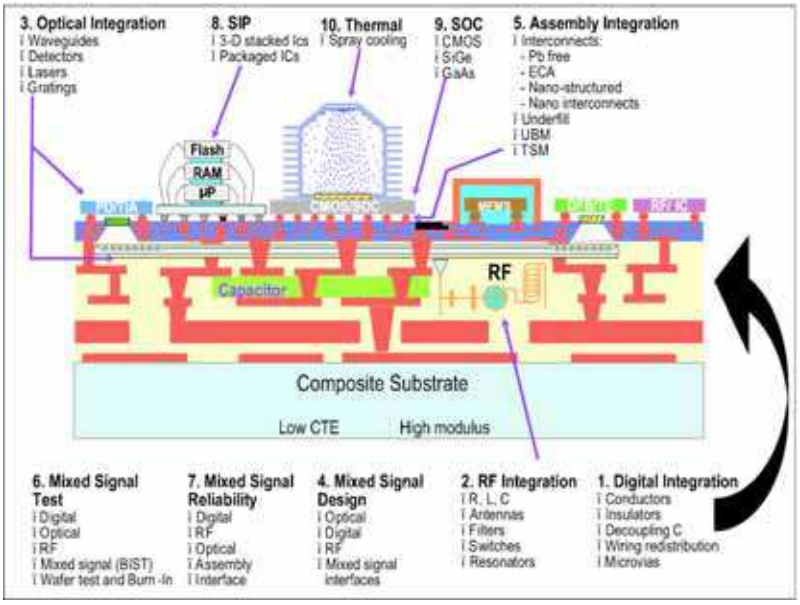



Fig. 2.3. SOP Integration by R.Tummala


Fig. 2.4. PillCam by GivenImaging

A perfect example of VSiP is the endoscopic capsule. GivenImaging developed the project PillCam [GivenImaging, 2005] for the human body exploration. This capsule is designed for examinations that could not be achieved by a classic analysis such the exploration of the small intestine. Once ingested, the PillCam begins to take pictures from inside the digestive system at a rate of two frames per second. For that it combines a CMOS imager, a light source, a unit of telecommunications to transmit the acquire picture and a battery. Unfortunately, the integrated 3D reconstruction into the form of SiP or SoC know a real delay. However, there is a lot of work along these lines we offer systems more and more miniaturized.


Fig. 2.5. SwissRanger SR3000 by CSEM

CSEM in Switzerland developed a 3D vision sensor based on the phase-measuring time-of-fly principle. The SwissRanger SR3000 [Oggier and al., 2006] looks like an autonomous module (fig2.) and is composed by three blocks: an integrated 3D-TOF image sensor, a 3D-camera electronics block with supplied power and a illumination blocks composed by 55 LEDS with a central wavelength of 850nm.

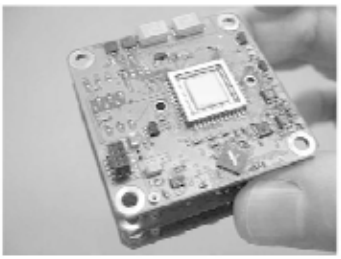


Fig. 2.6. Photograph of the PCB stack as implemented in the SR3000 camera

A light source emit a wave in the near-infrared which is intensity modulated with a few tens of MHz. When the wave reaches the scene, it is reflected. The result of the acquisition is a depth map on which we can apply the texture of the scene (fig.2.). Images processing are apply in the system in order to determinate in real time the quality of the information and if it is needed to apply a first filter or dedicated algorithms. The transmission protocol is the USB2.0 by default but it can be easily modified in order to by adaptable to owner utilization.

Fig. 2.7. Texture and Depth map of the SR3000 camera

An approach of passive stereoscopy has been presented by K.Konolige [Konolige]. And again it is not really going on the integrated system but on miniaturization. This project, the SRI Small Vision Module (SVM), is based on the principle of passive stereoscopy; it means that the information of distance comes from analysis of the scene from two different points of view. In this case the system embeds an algorithm of disparity on the picture with 64 levels (fig. 9). To realize this, the hardware consists of two CMOS 320x240 grayscale imagers and lenses, low-power A/D converters, a digital signal processor and a small flash memory for program storage. There are also a parallel port for the communication with the host machine and display.

This state of arts proofs that the microelectronic progresses permit to grow considerably an integration density on silicon. So it is possible from now on to make up systems which include multiples complexes functions using in the same time reduced silicon. There exist the numerous applications which can have benefit of this progress such as the endoscopies or sensors networks. Actually the application field of existing systems was too restricted because of their size or consummation but the integration in form of SoC, Sip or SoP gives a

new solution opening the way to emergent applications. The systems of integrated vision are increasingly developing for all kind of using like in medical area or even for the drone navigation. Unfortunately, the integrated stereoscopy is rarely today but it stays a challenge. In fact the 3D reconstruction systems are for the moment not more than the miniaturized or embedded systems. This chapter will follow up with the presentation of the adapted methods of reconstruction to the integration and an integrated 3D vision sensor able to answer on the real time restriction using perfectly controlled process of fabrication based on the standard Si technology.

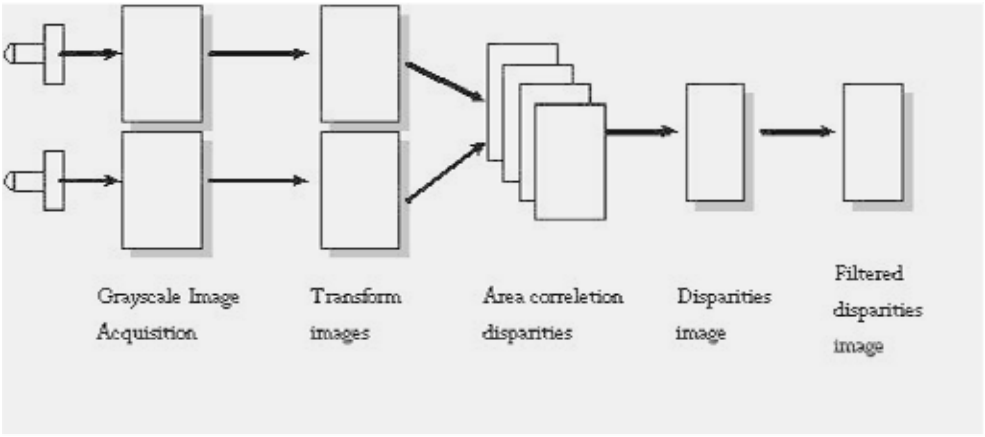



Fig. 2.8. Implemented images processing algorithm

## 3. Active vision and calibration

In this chapter we present the mathematical background for active vision system modeling.

### 3.1 System description

The active vision system is composed of an image sensor (a CMOS camera) and a structured pattern projector. The patterns studied here, are simple geometrical pattern obtained with a laser source and an optical diffraction head. This structure generates a repetitive geometrical pattern from a simple geometrical primitive. The primitives used in the pattern generator can be as simple as lines or dots.

The projector has to be simple in order to fit in the constrained size package of the embedded system. Thus complex pattern generators which can use DLP (as the ones used in video projectors) are excluded. The projected pattern considered hereafter is the simplest one, a mesh of regularly distributed points.

A multispectral approach with an IR (Infra–Red) projector and a multispectral (IR and visible images) CMOS sensor allows to grab simultaneously an image of the projected pattern and the scene texture. This is important for real time 3D generation but not only. Indeed, the multispectral approach helps to get easily a filtered image of the IR pattern.

The camera and the projector are considered rigidly fixed together fig.3.1 and the physical dimensions of the system do not change. This is important to allow off-line calibration of the vision system.

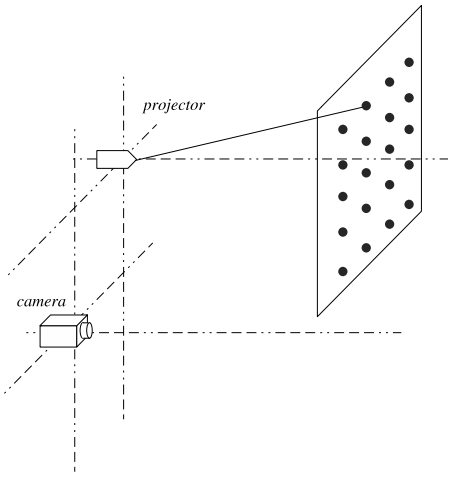



Fig. 3.1. Active stereoscopic system

The 3D reconstruction is achieved through triangulation fig.3.3. Each point of the projected pattern on the scene is the intersection of two lines:

- The line of sight, passing through the pattern point on the scene and its projection in the image plan. This line can be defined with the intrinsic and extrinsic parameters of the camera's pinhole model.
- The laser ray, starting from the projector center and passing through the chosen pattern point. This line can be defined by the projector parameters (position of the center and the angles between the rays).

The "a priori" knowledge of the stereoscopic system characteristics and the projected pattern structure, allows us to recover the distance to a projected spot of the pattern on the scene, from its image coordinate. The pattern simplicity implies correspondence ambiguities [Battle et al., 1998][Salvi et al., 2004][Salvi et al., 1998]. Indeed, we can not distinguish between laser spots in the scene from their shape or color. To solve the correspondence ambiguity problem, we use the epipolar constraint:
A laser impact can only belong, in the image, to the projection of the line supporting the laser ray. This projection represents the epipolar line [Faugeras, 1993]. We can thus simplify the correspondence search, limiting it to this line (one dimension search).

### 3.2 Calibration methods:

In this part we present two different calibration methods. Both allow us to obtain the correspondence between the position of the laser spots in the image and the distance to the object on which the laser is projected. We also obtain the epipolar lines equations to distinguish between the laser spots.
The first calibration method is described in [Marzani & Voisin, 2002]. Here we take several pictures of the pattern projected on a plan at different distances. As this method uses a big number of experimental points, it gives quite accurate models.
The second is analytic and has the advantage of only needing a unique picture shot to obtain the same results. We use here a calibration chart to calibrate the camera and the whole stereoscopic system and obtain thus a complete model describing the stereoscopic system.

### 3.2.1 1st method: Macroscopic method

**Epipolar model**

The pattern is projected on a plan surface parallel to the image plan. The projection plan is moved to different depth and each time a picture is taken. We thus obtain a set of points $(u_{jk}^n, v_{jk}^n)$ where j,k ∈ [1;7] are the position in the pattern mesh and $n$ references the image shot. These coordinates are expressed in pixel in the image. We can then obtain for each laser ray (j ,k), the epipolar line equations by fitting to a linear model

$$v = a \cdot u + b \tag{3.1}$$

so

$$\begin{pmatrix} v_{jk}^1 \\ v_{jk}^2 \\ \vdots \\ v_{jk}^n \\ \vdots \end{pmatrix} = \begin{pmatrix} u_{jk}^1 & 1 \\ u_{jk}^2 & 1 \\ \vdots & \vdots \\ u_{jk}^n & 1 \\ \vdots & \vdots \end{pmatrix} \begin{bmatrix} a_{jk} \\ b_{jk} \end{bmatrix}$$

or

$$V_{jk} = [U_{jk}\ 1] \begin{bmatrix} a_{jk} \\ b_{jk} \end{bmatrix}$$

The parameters *a* and *b* are calculated by a mean square approximation [Marzani & Voisin, 2002], as expressed in equation 3.2.

$$\begin{bmatrix} a_{jk} \\ b_{jk} \end{bmatrix} = \left( \begin{bmatrix} U_{jk}^T \\ 1 \end{bmatrix} [U_{jk}\ 1] \right)^{-1} \begin{bmatrix} U_{jk}^T \\ 1 \end{bmatrix} V_{jk} \tag{3.2}$$

**Depth model**

In addition to the epipolar lines, we have to express the relation between the position of a laser spot in the image and its distance to the stereoscopic system.

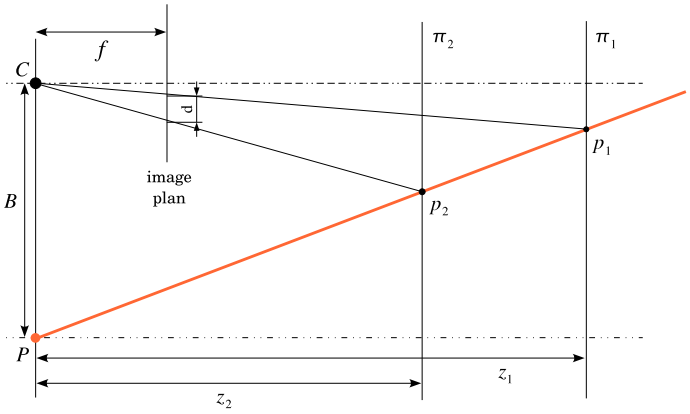



Fig. 3.2. Spot image movement vs. depth

If we consider a laser ray fig.3.2 projected on two different plans $\pi_1$ and $\pi_2$ located respectively at $z_1$ and $z_1$ from the projector/camera plan *CP* then, the trajectory $d$ of the spot impact in the image (this displacement is constrained to the epipolar line corresponding to the laser ray). Considering the two triangles $CPp_1$ and $CPp_2$, we can express $d$ as:

$$d = B\left[\left(\frac{z_1 - f}{z_1}\right) - \left(\frac{z_2 - f}{z_2}\right)\right] = Bf\frac{(z_1 - z_2)}{z_1 z_2} \tag{3.3}$$

Where $B$ is the stereoscopic base and $f$ focal length of the camera and $d$ is the shift between the image positions $(u_1, v_1)$ and $(u_2, v_2)$ on the epipolar line.

$$d = \sqrt{(u_1 - u_2)^2 + (v_1 - v_2)^2}$$

Given the epipolar line equation 3.1 we can have an expression depending only on one coordinate:

$$d = \sqrt{1 + a^2}\,(u_1 - u_2) \tag{3.4}$$

From the lasts equations 3.3, 3.4 and by choosing the appropriate referential origins, we can express the depth as a hyperbolic function of the $u$ image coordinate, equation 3.5.

$$z = \frac{1}{\alpha u + \beta} \tag{3.5}$$

Where the $\alpha$ and $\beta$ parameters are also calculated using a mean square method.

$$\begin{pmatrix} 1/z^1 \\ 1/z^2 \\ \vdots \\ 1/z^n \\ \vdots \end{pmatrix} = \begin{pmatrix} u^1_{pk} & 1 \\ u^2_{pk} & 1 \\ \vdots & \vdots \\ u^n_{pk} & 1 \\ \vdots & \vdots \end{pmatrix} \begin{bmatrix} \alpha_{pk} \\ \beta_{pk} \end{bmatrix}$$

or

$$\hat{Z} = [U_{pk}\ 1] \begin{bmatrix} \alpha_{pk} \\ \beta_{pk} \end{bmatrix}$$

and so

$$\begin{bmatrix} \alpha_{pk} \\ \beta_{pk} \end{bmatrix} = \left( \begin{bmatrix} U^T_{pk} \\ 1 \end{bmatrix} [U_{pk}\ 1] \right)^{-1} \begin{bmatrix} U^T_{pk} \\ 1 \end{bmatrix} \hat{Z}$$

where $\hat{Z}$ is the inverse depths vector.

### 3.2.1 2nd method: analytic method

We aim here to obtain a more complete characterization of the whole stereoscopic system from an image in a single snapshot.

**Camera calibration**

The projective model of the camera (considered ideal), also call pinhole model, links the coordinates of a point in the "world" referential to the image [Faugeras, 1993][Weng et al., 1990][Battle et al., 1998][ Horaud & Monga, 1995]. It describes the relative position between the "world" and the image referential and the projection process.

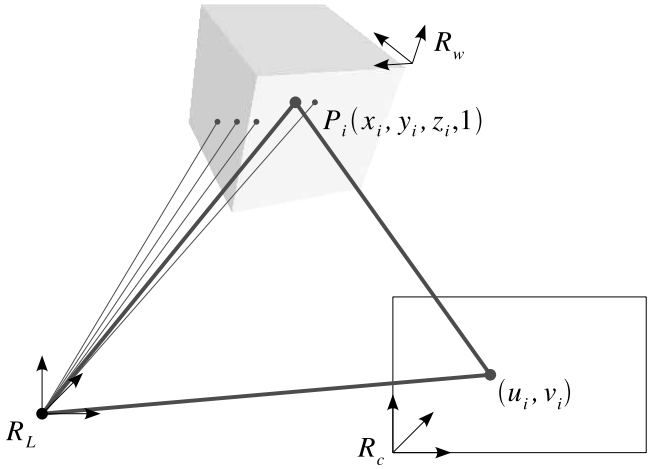



Fig. 3.3. Camera and World referential

For commodity reasons, this model is expressed in homogeneous coordinate. Thus translations and rotations are represented by matrix to vector multiplications.

The model can be split into two parts:

**Extrinsic model**: Represents the relation between the chosen world referential $R_w$ and the camera referential $R_c$ (see fig.3.3). It expresses all the translations and rotations between the two referential.

$$R_{3\times3} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \text{ and } T_{1\times3} = \begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix}$$

are respectively these rotations and the translation matrix.

The transformation can be expressed in homogeneous coordinates as:

$$\begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}_{R_c} = \begin{bmatrix} R_{3\times3} & T_{3\times1} \\ 0_{1\times3} & 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}_{R_w}$$

**Intrinsic model**: Describes the projective imaging transformation.

It links a point coordinate expressed in the camera referential with the image coordinate expressed in pixels. Figure 3.4 shows the projection process of a point $P_c(x, y, z)^T_{Rc}$ expressed in the camera coordinates. To simplify the representation only $x$ and $z$ axis are represented.

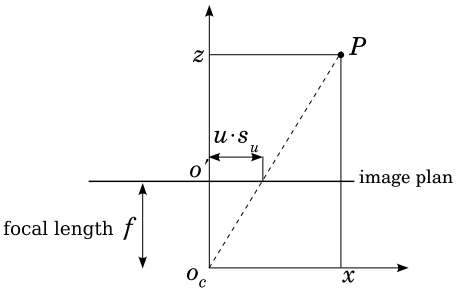



Fig. 3.4. Camera projection

$O_c$ is the camera optical center and $O'$ the projection of this center on the image plan which image coordinates are $u_0$, $v_0$. ($f$) the focal length and ($s_u \times s_v$) the pixel size.
We can express the image coordinate of the point:

$$\begin{cases} z \cdot (u - u_0) = x \cdot \frac{f}{s_u} \\ z \cdot (v - v_0) = y \cdot \frac{f}{s_v} \end{cases}$$

Using homogeneous coordinates we can express this last equation as:

$$\begin{bmatrix} \lambda \cdot u \\ \lambda \cdot v \\ \lambda \end{bmatrix} = I \cdot \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}_{R_c} = \begin{bmatrix} f/s_u & 0 & u_0 & 0 \\ 0 & f/s_v & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}_{R_c}$$

Where $I$ is the intrinsic matrix, and $\lambda = z$ is the projection proportionality reduction factor.
The intrinsic matrix $I$ has a null column and so is irreversible which means that we can not retrieve the depth from the sole image coordinate.
The global camera model is thus a matrix to vector multiplication connecting a point coordinates expressed in the "world referential" to the image coordinates expressed in pixels.

$$\begin{bmatrix} \lambda \cdot u \\ \lambda \cdot v \\ \lambda \end{bmatrix} = M \cdot \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}_{R_w} = I \cdot \begin{bmatrix} R_{3\times3} & T_{3\times1} \\ 0_{1\times3} & 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}_{R_w} \tag{3.6}$$

Where

$$M = \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{bmatrix}$$

is the fundamental matrix.

We can rearrange equation 3.6 as:

$$\begin{cases} u = \dfrac{m_{11}\cdot x + m_{12}\cdot y + m_{13}\cdot z + m_{14}}{m_{31}\cdot x + m_{32}\cdot y + m_{33}\cdot z + m_{34}} \\ v = \dfrac{m_{21}\cdot x + m_{22}\cdot y + m_{23}\cdot z + m_{24}}{m_{31}\cdot x + m_{32}\cdot y + m_{33}\cdot z + m_{34}} \end{cases} \tag{3.7}$$

To obtain the necessary points for the fundamental matrix parameters estimation, we have design a new calibration 3D pattern fig.3.5 .

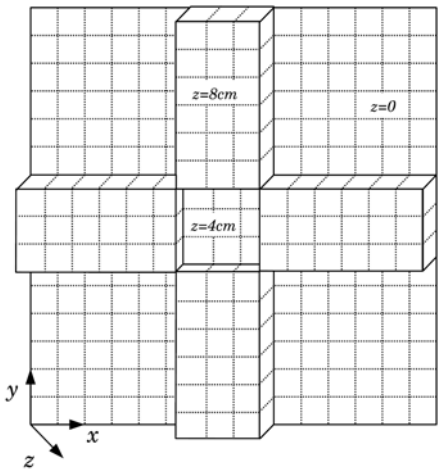



Fig. 3.5. Calibration chart referential

An automatic corner extraction, using Harris filter, from an image of the 3D chart, gives us enough points (184) for a robust estimation [Battle et al., 1998].
Using this model we can obtain the line of sight from any point in the image and especially for the projected pattern points. Thus, for each pattern point $i$, we determine the line of sight as an intersection between two plans as:

$$\begin{cases} \Gamma_{11}x_i + \Gamma_{12}y_i + \Gamma_{13}z_i + \Gamma_{14} = 0 \\ \Gamma_{21}x_i + \Gamma_{22}y_i + \Gamma_{23}z_i + \Gamma_{24} = 0 \end{cases} \tag{3.8}$$

$$\text{with} \quad \begin{cases} \Gamma_{11} = (m_{11} - u_i \cdot m_{31}) \\ \Gamma_{12} = (m_{12} - u_i \cdot m_{32}) \\ \Gamma_{13} = (m_{13} - u_i \cdot m_{33}) \\ \Gamma_{14} = (m_{14} - u_i \cdot m_{34}) \\ \Gamma_{21} = (m_{21} - v_i \cdot m_{31}) \\ \Gamma_{22} = (m_{22} - v_i \cdot m_{32}) \\ \Gamma_{23} = (m_{23} - v_i \cdot m_{33}) \\ \Gamma_{24} = (m_{24} - v_i \cdot m_{34}) \end{cases}$$

**Laser projector characterization:**
To validate experimentally the method, we have used a 7x7 laser matrix pattern. Each laser ray can be considered as the intersection of two supporting plans fig.3.6. Thus, we can define the 49 laser rays by the parameters of 7 horizontal plans ($H_j$) and 7 vertical plans ($V_k$).

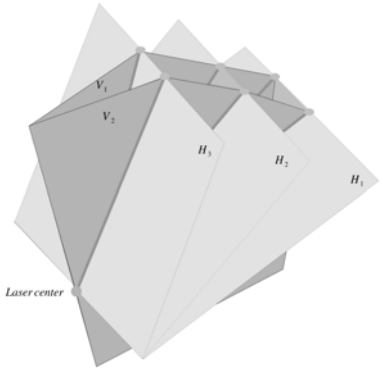



Fig. 3.6. Laser projector model

$$\begin{cases} x_i \cdot h_{1j} + y_i \cdot h_{2j} = z_i + h_{3j} \\ x_i \cdot v_{1k} + y_i \cdot v_{2k} = z_i + v_{3k} \end{cases} \quad with\ j,k \in [1;7] \tag{3.9}$$

To determine the plan parameters, we use an IR image of the laser pattern projected on the 3D calibration chart The chart shape helps us to get easily the $z_i$ coordinate of each spot.
The $x_i$ and $y_i$ coordinates are obtained from the image coordinate $z_i$ and the fundamental matrix using (3.8). Arranging the equation, we obtain thus the $x_i$ and $y_i$ world coordinate of each point (3.10).

$$\begin{bmatrix} x_i \\ y_i \end{bmatrix} = M_{xy}^{-1} \cdot M_{xyz} \tag{3.10}$$

where,

$$M_{xy} = \begin{bmatrix} (m_{11} - a_1 \cdot m_{31}) & (m_{12} - a_1 \cdot m_{32}) \\ (m_{21} - a_2 \cdot m_{31}) & (m_{22} - a_2 \cdot m_{32}) \end{bmatrix}$$

and

$$M_{xyz} = \begin{bmatrix} z_i \cdot (m_{13} - a_1 \cdot m_{33}) + (m_{14} - a_1 \cdot m_{34}) \\ z_i \cdot (m_{23} - a_2 \cdot m_{33}) + (m_{24} - a_2 \cdot m_{34}) \end{bmatrix}$$

Furthermore, we can identify 7 points verifying each plan equation. So we can write for each horizontal plan:

$$\begin{pmatrix} x_{1j} & y_{1j} & 1 \\ \vdots & \vdots & \vdots \\ x_{ij} & y_{ij} & 1 \\ \vdots & \vdots & \vdots \\ x_{7j} & y_{7j} & 1 \end{pmatrix} \cdot \begin{bmatrix} h_{1j} \\ h_{2j} \\ h_{3j} \end{bmatrix} = - \begin{pmatrix} z_{1j} \\ \vdots \\ z_{ij} \\ \vdots \\ z_{7j} \end{pmatrix}$$

or

$$K_k \cdot H_j = -Z_j$$

The horizontal plans parameters can be finally obtained by:

$$H_j = -(K_h^T \cdot K_h^{-1})^{-1} \cdot K_h^T \cdot Z_j$$

The same result can be obtained for the vertical plans:

$$V_k = -(K_v^T \cdot K_v^{-1})^{-1} \cdot K_v^T \cdot Z_k$$

The determination of the plans parameters allows us to determine the projector center in the world referential.
Indeed, this point is the intersection of all the horizontal and vertical plans, and verifies the equation 3.11.

$$\begin{pmatrix} h_{11} & h_{21} & 1 \\ \vdots & \vdots & \vdots \\ h_{1T} & h_{2T} & 1 \\ v_{11} & v_{21} & 1 \\ \vdots & \vdots & \vdots \\ v_{1T} & v_{2T} & 1 \end{pmatrix} \cdot \begin{bmatrix} x_{laser} \\ y_{laser} \\ z_{laser} \end{bmatrix} = - \begin{pmatrix} h_{31} \\ \vdots \\ h_{3T} \\ v_{31} \\ \vdots \\ v_{3T} \end{pmatrix} \tag{3.11}$$

**Depth recovery**

Once we have characterized the whole stereoscopic system (camera perspective model and the projector parameters), we can recover the coordinate of a laser spot (i) projected on any object from its image coordinate and the laser index (jk).
Starting from the laser ray equation 3.9 we can express the $x_i$ and $y_i$ coordinates as:

$$\begin{bmatrix} x_i \\ y_i \end{bmatrix} = \begin{bmatrix} h_{1j} & h_{2j} \\ v_{1k} & v_{2k} \end{bmatrix}^{-1} \cdot \begin{bmatrix} z_i + h_{3j} \\ z_i + v_{3k} \end{bmatrix} = \begin{bmatrix} \frac{v_{2k}(z_i + h_{3j}) - h_{2j}(z_i + v_{3k})}{h_{1j} \cdot v_{2k} - h_{2j} \cdot v_{1k}} \\ \frac{h_{1j}(z_i + v_{3k}) - v_{1j}(z_i + h_{3j})}{h_{1j} \cdot v_{2k} - h_{2j} \cdot v_{1k}} \end{bmatrix}$$

$$\text{thus} \quad \begin{bmatrix} x_i \\ y_i \end{bmatrix} = \begin{bmatrix} l_{1jk} \cdot z_i + l_{2jk} \\ l_{3jk} \cdot z_i + l_{4jk} \end{bmatrix} \tag{3.12}$$

with

$$l_{1jk} = \frac{v_{2k} - h_{2j}}{h_{1j} \cdot v_{2k} - h_{2j} \cdot v_{1k}} = \frac{v_{2k} - h_{2j}}{\Delta}$$

$$l_{2jk} = \frac{v_{2k} \cdot h_{3j} - h_{2j} \cdot v_{3j}}{h_{1j} \cdot v_{2k} - h_{2j} \cdot v_{1k}} = \frac{v_{2k} \cdot h_{3j} - h_{2j} \cdot v_{3j}}{\Delta}$$

$$l_{3jk} = \frac{h_{1j} - v_{1k}}{h_{1j} \cdot v_{2k} - h_{2j} \cdot v_{1k}} = \frac{h_{1j} - v_{1j}}{\Delta}$$

$$L_{4jk} = \frac{v_{1k} \cdot h_{3j} - h_{1j} \cdot v_{3k}}{h_{1j} \cdot v_{2k} - h_{2j} \cdot v_{1k}} = \frac{v_{1k} \cdot h_{3j} - h_{1j} \cdot v_{3k}}{\Delta}$$

Where

$$\Delta = h_{1j} \cdot v_{2k} - h_{2j} \cdot v_{1k}.$$

By replacing 3.12 in 3.7 we obtain the image coordinate ($u_i$, $v_i$) versus the depth $z_i$.

$$u_i = \frac{A^1_{jk} \cdot z_i + A^2_{jk}}{A^3_{jk} \cdot z_i + A^4_{jk}}$$

$$v_i = \frac{A^5_{jk} \cdot z_i + A^6_{jk}}{A^3_{jk} \cdot z_i + A^4_{jk}}$$

where

$$\begin{cases} A^1_{jk} = m_{11} \cdot l_{1jk} + m_{12} \cdot l_{3jk} + m_{13} \\ A^2_{jk} = m_{11} \cdot l_{2jk} + m_{12} \cdot l_{4jk} + m_{14} \\ A^3_{jk} = m_{31} \cdot l_{1jk} + m_{32} \cdot l_{3jk} + m_{33} \\ A^4_{jk} = m_{31} \cdot l_{2jk} + m_{32} \cdot l_{4jk} + m_{34} \\ A^5_{jk} = m_{21} \cdot l_{1jk} + m_{22} \cdot l_{3jk} + m_{23} \\ A^6_{jk} = m_{21} \cdot l_{2jk} + m_{22} \cdot l_{4jk} + m_{24} \end{cases}$$

Hence, by eliminating $z_i$ from the last equations we obtain the epipolar line equation:

$$[A^5_{jk}A^4_{jk} - A^3_{jk}A^6_{jk}] \cdot u_i + [A^3_{jk}A^2_{jk} - A^1_{jk}A^4_{jk}] \cdot v_i = [A^2_{jk}A^5_{jk} - A^1_{jk}A^6_{jk}]$$

We identify the epipolar line parameters as in equation 3.1:

$$\begin{cases} a_{jk} = -\frac{[A^5_{jk}A^4_{jk} - A^3_{jk}A^6_{jk}]}{[A^5_{jk}A^2_{jk} - A^1_{jk}A^6_{jk}]} \\ b_{jk} = \frac{[A^1_{jk}A^5_{jk} - A^1_{jk}A^6_{jk}]}{[A^3_{jk}A^2_{jk} - A^1_{jk}A^4_{jk}]} \end{cases}$$

We can also express the $z_i$ coordinate as a function of one of the image coordinate:

$$z_i = \frac{A^4_{jk} \cdot u_i - A^2_{jk}}{A^1_{jk} - A^3_{jk} \cdot u_i}$$

**Particular case:**

If there is no rotation between the image and the world referential and the $z$ axis translation is null (as represented in fig.3.2). then, we have:

$$R_{3\times3} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

and

$$T_{1x3} = \begin{bmatrix} t_1 \\ t_2 \\ 0 \end{bmatrix}$$

the fundamental matrix will be

$$M = \begin{bmatrix} f/s_u & 0 & u_0 & t_1 f/s_u \\ 0 & f/s_v & v_0 & t_2 f/s_v \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

with $m_{31}$, $m_{32}$ and $m_{34}$ equal to zero and so $A^4_{jk} = 0$.
Thus, by choosing a good physical configuration a relation, between the world coordinate $z$ and the image coordinate $u$, equivalent to what expressed in 3.5.

## 4. Cyclope

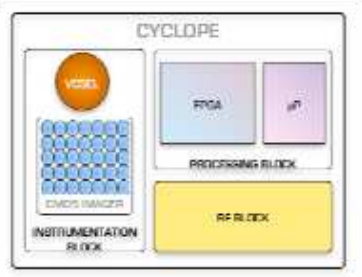



Fig. 4.1. Synoptic of Cyclope

Cyclope is an integrated 3D vision system sensor based on active vision. It is composed of three essential parts:

- An instrumentation block: for image grabbing and the generation of the structured light pattern.
- A processing block: for control and data processing.
- A RF block: for the transmission of the results and the dynamic OTA (Over The Air) reconfiguration.

These parts are realized in different technologies: CMOS for the image sensor and the processing units; GaAs for the pattern projector and RF–CMOS for the communication unit. The monolithic integration of those elements is not feasible, this is why the development of an integrated "SIP" (System In Package) is actually the best solution to overcome the technological constraints and realize a chip scale package. This solution is used in several embedded sensors as The Human++ platform [Gyselinckx et al., 2005] or Smart Dust [Warneke et al., 2001] where digital processing has to cohabit with optical and wireless communication. In Cyclope, we have chosen the realization of an integrated system based on active vision. This choice is the consequence of two facts. The first is the real time processing that we would like to obtain. In active vision, a known pattern is projected on a scene, this projection belong a specific line denoted as epipolar line. This property is very interesting in real time processing, because it could reduce the complex problem of the

computation on the z coordinate to a simple problem of line computation. The second fact is that active vision simplifies the matching between a point and is projection, because the only points that we want to match, is the pattern points.

**A processing block**

As we define in the last part, we have a set of parameters for the epipolar and depth models. Those parameters are used on run time to make the point matching (identify the original position of a pattern point from its image) and calculate the depth using the appropriate parameters.

The architecture is divided into two principal parts:

- Pre-processing unit for low level image processing and feature extraction from the IR image.
- 3D unit for point matching and depth calculation.

Figure 4.2 shows the global processing architecture. A dual-port memory is used for image storage allowing asynchronous image acquisitions. The processing implemented in the Pre-processing unit is thresholding, segmentation and spot center calculation. A FIFO allows the communication between the two units and a final storage FIFO allows communication toward an external UART.

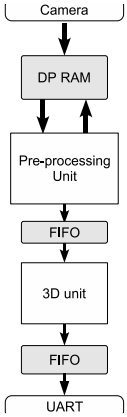



Fig. 4.2. Global processing architecture

### 4.1.1 The 3D unit

In this part we present the 3D extraction method for on line processing. To achieve this purpose, we have designed a parallel digital processing unit fig.4.3.

After a pre-processing step, where the laser spot center is estimated from the IR image, the coordinate of each point are directed to the processing unit where the (z) coordinate is calculated.

Starting from the point abscissa (u) we calculate its estimated ordinate (~v) if it belongs to an epipolar line. We compare this estimation with the true ordinate (v).

These operations are made for all the epipolar line simultaneously. After thresholding the encoder returns the index of the corresponding epipolar line.

The next step is to calculate the z coordinate from the u coordinate and the appropriate depth model parameters.

These computation blocs are synchronous and pipelined, allowing thus, high processing rates.

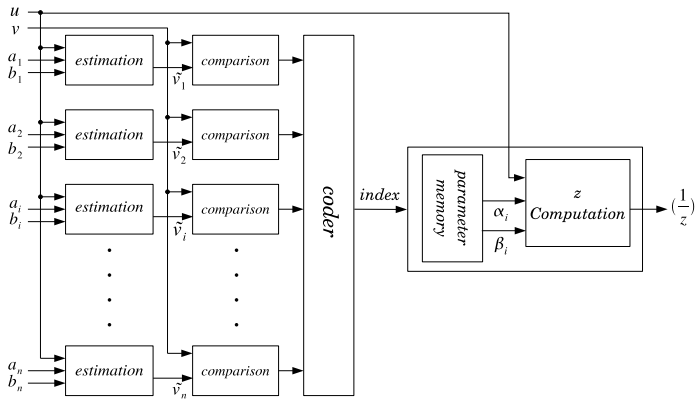



Fig. 4.3. 3D unit

### 4.1.2 Estimation bloc:

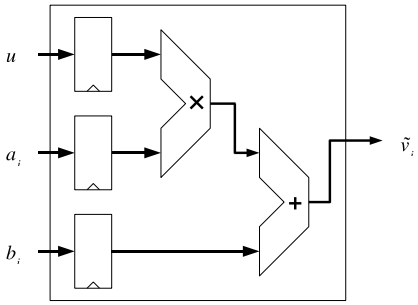


Fig. 4.4. Estimation bloc

In this bloc the estimated ordinate is calculated ~v=a.u+b . The (a,b) parameters are loaded from memory.

### 4.1.3 Comparison bloc:

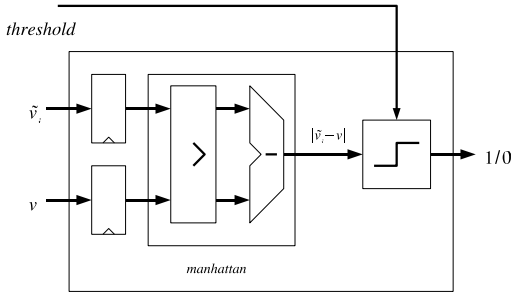



Fig. 4.5. Comparison bloc

In this bloc the absolute value of the difference between the ordinate v and its estimation $\tilde{v}$ is calculated. This difference is then thresholded.
The thresholding avoids a resource consuming sort stage. The threshold was a priori chosen as half the minimum distance between two consecutive epipolar lines. The threshold can be adjusted for each comparison bloc.
This bloc returns a '1' result if the distance is underneath the threshold.

#### 4.1.4 Encoding bloc:

If the comparison blocs returns a unique '1' result, then the encoder returns the corresponding epipolar line index.
If no comparison bloc returns a 'true' result, the point is irrelevant and considered as picture noise.
If more then one comparison bloc returns '1' then we consider that we have a correspondence error and a flag is set. The selected index is then carried to the next stage where the z coordinate is calculated. It allows the selection of the good parameters to the depth model.
We have chosen to calculate (1/z) rather than z to have a simpler computation unit. This computation bloc is then identical to the estimation bloc.

### 4.2 The instrumentation block

#### 4.2.1 An energetic approach

Having opted for a laser pattern which give out in the near infrared spectrum in order to not misrepresent the scene, many classic solution were possible. We must acquire in the same time an image of the texture therefore in the visible spectrum, and an image of the projected pattern:

- the utilization of a BAYER type optic filter (VI/IR) .
- the conception of a CMOS imager which used a semiconductor technology with a near infrared spectral response [Pei et al., 2003] or by changing the depth and the nature of the junctions [David Starikov & Bensaoula, 2004]
- made the separation of the spectra by electronic or image processing.

Unfortunately none of these options could not completely satisfy us. Indeed, we wanted a solution easily implemented by simplifying the fabrication process and thus stay in a standard Silicon technology without any optic filter, adaptable to already existing systems by providing them with a 3D information without major changes, and finally be able to assure a classic video rate of 25 images per second.
It is in this context that we opted for an original acquisition approach based on the energetic's gap between the scene and the pattern and not based on the wave length.
The idea is to project a pattern much more energetic than the texture could be and to modify the integration time of the imager in order to only acquire the projected pattern. And if is needed, we can also realized an image processing like a binarization.

#### 4.2.2 Image acquisition processing:

To allow real–time acquisition of both pattern and texture, we have developed a multi-spectral 64x64 pixels CMOS imager (fig. 4.6). This sensor has programmable light integration and shutter time to allow dynamic change. The projector pulses periodically on the scene an energetic pattern. An image acquisition with a short integration time allows to

grab the image of the pattern without the background texture. A second image acquisition with a longer integration allows to grab the texture when the projector is off. The figure 4.7 shows the sequential scheduling of the images acquisition. To reach a video frame rate of 25 images/s this acquisition sequence must be done in less than 40 ms. The global acquisition time is given in equation 4.1 where $T_{rst}$ is the reset time, $T_{rd}$ is the time needed to read the entire image and $T_{intVI}$ $T_{intIR}$ are respectively the integration time for both visible and IR image.

$$T_{total}=2\cdot T_{rst}+2\cdot T_{rd}+T_{intVI}+T_{intIR} \tag{4.1}$$

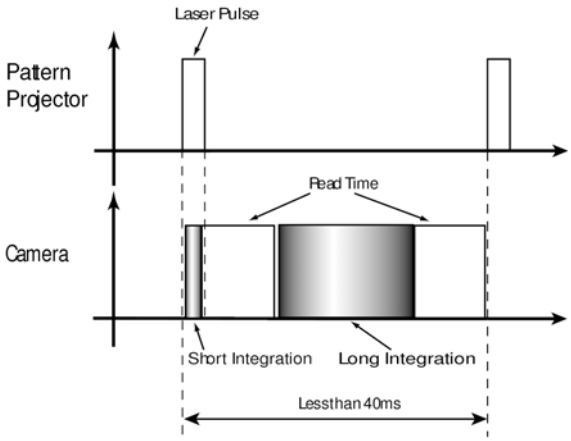



Fig. 4.7. Acquisition sequence

## 5. Conclusion

In this chapter we saw that the issues of the integrated stereoscopy are really important. Many of emergent domains such as the 3D endoscopy or network sensors for the 3D mapping directly depend on the advance of the integrated 3D vision because the existing solutions is limiting by their application domain, their size or autonomy and there are not suitable to an integrated use. The integrated 3D vision is able to give us a solution where it exist heavy material constraint but we need to develop special architectures and methods in order to have an efficiency solution.

In order to have a complete view of this challenge, we proposed an integrated active stereoscopic vision sensor design for emergent domain and the mathematical model that we developed in order to be used in such applications. Based on this model we have designed a processing architecture allowing a high integration level.

We also proposed an original acquisition method based on the energetic's gap between the texture and the projected pattern allowing an easily spectrum discrimination.

The next step to this work is the chip level integration of both the image sensor and the pattern projector. Evaluate the power consumption of the pulsed laser projector considering the optical efficiency of the diffraction head.

## 6. References

[Battle et al., 1998] J. Battle, E. Mouaddib, and J. Salvi. Recent progress in codded structurred light as a technique to solve the correspondance problem: a survey. Pattern Recognition, 31(7) :963-982, 1998.

[Belhumeur, 1993] P.N. Belhumeur. A binocular stereo algorithm for reconstructing sloping, creased, and broken surfaces in the presence of half-occlusion. In Fourth International Conference on Computer Vision, 1993.

[Chen and Li, 2003] S.Y. Chen and Y.F. Li. A 3d vision system using unique color encoding. In Internanional conference on robotics, intelligent systems and signal processing, 2003.

[Chichyang and Zheng, 1998] Chen Chichyang and Y.F. Zheng. Passive and active stereo vision for smooth surface detection of deformed plates. IEEE Transactions on Pattern Analysis and Machine Intelligence, 7 :62-81, 1998.

[Cochran and Medioni, 1992] S.D. Cochran and G. Medioni. 3-d surface description from binocular stereo. IEEE Transactions on Pattern Analysis and Machine Intelligence, 14 :981-994, 1992.

[David Starikov and Bensaoula, 2004] Rajeev Pillai David Starikov, Chris Boney and Abdelhak Bensaoula. Dual-band uv/ir optical sensors for fire and flame detection and target recognition. Senson for Industry Conference, 2004.

[Faugeras, 1993] O. Faugeras. Three-Dimensional Computer Vision, a Geometric Viewpoint. MIT Press, 1993.

[GivenImaging, 2005] GivenImaging. Given Diagnostic System, The Platform for PillCam Endoscopy. 2005.

[Gokturk et al., 2004] S.B. Gokturk, H. Yalcin, and C. Bamji. A time-of-flight depth sensor - system description, issues and solutions. In Conference on Computer Vision and Pattern Recognition Workshop, 2004. 33

[Gruss et al., 1991] A. Gruss, L.R. Carley, and T. Kanade. Integrated sensor and range-finding analog signal processor. IEEE Journal of Solid-State Circuits, 26 :184-191, 1991.

[Gyselinckx et al., 2005] B. Gyselinckx, C. Van Hoof, J. Ryckaert, R.F. Yazicioglu, P. Fiorini, and V. Leonov. Human++ : autonomous wireless sensors for body area networks. In IEEE 2005 Custom Integrated Circuits Conference, 2005.

[Horaud and Monga, 1995] R. Horaud and O. Monga. Vision par ordinateur. Edition Hermès, 1995.

[Hyun and Gerhardt, 1994] K. Hyun and L.A. Gerhardt. The use of laser structured light for 3d surface measurement and inspection. In Proceedings of the Fourth International Conference on Computer Integrated Manufac-turing and Automation Technology, 1994.

[Konolige, ] K. Konolige. Small vision system : Hardware and implementation.

[Lange and Seitz, 2001] R. Lange and P. Seitz. Solid-state time-of-flight range camera. IEEE Journal ofQuantum Electronics, 37 :390-397, 2001.

[Marr and Poggio, 1979] D. Marr and T. Poggio. A computational theory of human stereo vision. In Proceedings of the Royal Society of London B, volume 204, pages 301-328, 1979.

[Marzani and Voisin, 2002] F. Marzani and Y. Voisin. Calibration of a three-dimentionnal reconstruction system using a structured light source. Optical Engineering, 41 :484-492, 2002.

[Moini, 1999] A. Moini. Vision Chips. Kluwer Academic Publishers, 1999.

[Oggier et al., 2006] T. Oggier, B.Büttgen, F.lustenberger, G.Becker, B. Rüegg, and A.Hodac. Swissranger sr3000 and first experiences based on miniaturized 3d-tof cameras. 2006.

[Pei et al., 2003] Z. Pei, L.S. Lai, H.P. Hwang, Y.T. Tseng, and M.-J. Tsai C.S. Liang. Si1 xgex=si multi-quantum well phototransistor for near-infrared operation. Physical, 2003.

[Pinna, 2003] A. Pinna. Conception d'une rétine connexionniste : du capteur au système de vision sur puce. PhD thesis, Paris 6 - UPMC, 2003.

[Salvi et al., 1998] J. Salvi, J. Batlle, and E. Mouaddib. A robust-coded pattern projection for dynamic 3d scene measurement. Pattern Recognition Letter, 19 :1055-1065, 1998.

[Salvi et al., 2004] J. Salvi, J. Pagès, and J. Batlle. Pattern codification strategies in structured light systems. Pattern Recognition, 37 :827-849, 2004.

[Song and Wang, 2006] LiMei Song and DaNvWang. A novel grating matching method for 3d reconstruction. NDT & E International, 39 :282-288, 2006.

[Song et al., 2003] Yong-Ha Song, S.G. Kim, K.J Rhee, and T.S Kim. A study of considering the reliability issue on asic/memoty integration by sip (system-in-package) technologie. Microelectronics Reliability, 2003.

[Stooke, 1991] J. Stooke. Cartographiy of asteroides and comet nuclei from low resolution data. Asteroids, Comets, Meteors, pages 583-586, 1991.

[Szeliski, 2000] R. Szeliski. Scene reconstruction from multiple cameras. In International Conference on Image Processing, 2000.

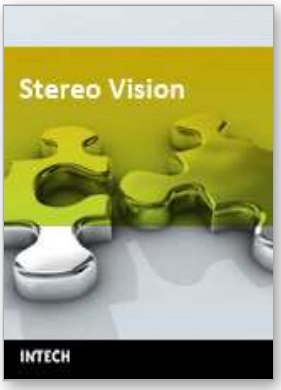
[Tummala, 2005] R.R Tummala. Packaging : past, present and future. In International Conference on Electronic Packaging Technology, 2005.

[Valkenburg and McIvor, 1998] R.J. Valkenburg and A.M. McIvor. Accurate 3d measurement using a structured light system. Image and Vision Computing, 16 :99-110, 1998.

[Vézien, 1995] Jean-Marc Vézien. Techniques de reconstruction globale par analyse de paire d'images stéréoscopiques. PhD thesis, Paris 7, 1995.

[Warneke et al., 2001] B. Warneke, M. Last, B. Liebowitz, and K.S.J Pister. Smart dust. communicating with a cubic milimeter computer. Computer, 34 :44 - 51, Jan 2001.

[Weng et al., 1990] J. Weng, P. Cohen, and M. Herniou. Calibration of stereo cameras using a non-linear distortion model. In Pattern Recognition, 1990. Proceedings., 10th International Conference on, volume i 16-21 Jun 1990, pages 246-253 vol.1, 1990.




**Stereo Vision**

Edited by Asim Bhatti

ISBN 978-953-7619-22-0
Hard cover, 372 pages
**Publisher** InTech
**Published online** 01, November, 2008
**Published in print edition** November, 2008

The book comprehensively covers almost all aspects of stereo vision. In addition reader can find topics from defining knowledge gaps to the state of the art algorithms as well as current application trends of stereo vision to the development of intelligent hardware modules and smart cameras. It would not be an exaggeration if this book is considered to be one of the most comprehensive books published in reference to the current research in the field of stereo vision. Research topics covered in this book makes it equally essential and important for students and early career researchers as well as senior academics linked with computer vision.

**How to reference**

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Kolar Anthony, Romain Olivier, Graba Tarik, Ea Thomas and Granado Bertrand (2008). The Integrated Active Stereoscopic Vision Theory, Integration and Application, Stereo Vision, Asim Bhatti (Ed.), ISBN: 978-953-7619-22-0, InTech, Available from: http://www.intechopen.com/books/stereo_vision/the_integrated_active_stereoscopic_vision_theory__integration_and_application

INTECH
open science | open minds

**InTech Europe**
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

**InTech China**
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821