A methodology to develop Computer Vision systems in Civil Engineering: Applications in material testing and fish tracking

Autor/a: Álvaro Rodríguez Tajes

Doctoral Thesis UDC / 2014

Director/a: Juan Ramón Rabuñal Dopico

RD 778/1998: Department of Comunication and Information Technologies

UNIVERSIDADE DA CORUÑA

A quén corresponda

All that we see or seem is but a dream within a dream

Edgar Allan Poe

The realization of this PhD. has been a long journey, and some people had supported me both consciously and unconsciously. These people (family, friends, colleagues and bartenders) are too many to be named here. To you, my sincere gratitude.

The help of Juan Luis and Carlos was really appreciated (but I am too lazy to invite you to a beer).

The author also wishes to express his condolences to Maria and Miki, who wasted a significant part of their life watching fishes.

Álvaro Rodríguez Tajes

ACKNOWLEDGMENTS

The research carried out in this thesis would not have been possible without the collaboration and support of RNASA lab (Dept. of Information and Communications Technologies, Faculty of Informatics), the GEAMA group (Water and Environmental Engineering Group, CITEEC) and the GCONS group (Construction Group, CITEEC) of the University of A Coruña.

The CITIC (Investigation Center in Information and Communication Technologies), the CITEEC (Center of Technological Innovations in Construction and Civil Engineering) and the CEH (Center for Hydrographic Studies) of CEDEX (Center for Studies and Experimentation of Public Works) had supported this research with their staff and facilities.

This research was also supported by Dirección Xeral de Investigación, Desenvolvemento e Innovación of Xunta de Galicia (Ref. 10MDS014CT and Ref. 08TMT005CT), Ministerio de Economía y Competitividad (Ref. CGL2012-34688 PCT-380000-2007-3) and it was partially funded with FEDER funds, and by Centro de Estudios Hidrográficos (CEH) of CEDEX.

ABSTRACT

Computer Vision provides a new and promising approach to Civil Engineering, where it is extremely important to measure with accuracy real world processes.

However, Computer Vision is a broad field, involving several techniques and topics, and the task of defining a systematic development approach is problematic.

In this thesis a new methodology is carried out to develop these systems attending to the special characteristics and requirements of Civil Engineering. Following this methodology, two systems were developed:

A system to measure displacements from real images of material surfaces taken during strength tests. This technique solves the limitation of current physical sensors, which interfere with the assay and which are limited to obtaining measurements in a single point of the material and in a single direction of the movement.

A system to measure the trajectory of fishes in vertical slot fishways, whose purpose is to solve current lacks in the design of fishways by providing information of fish behavior.

These applications represent significant contributions to the field and show that the defined and implemented methodology provides a systematic and reliable framework to develop a Computer Vision system in Civil Engineering.

RESUMEN

La Visión Artificial proporciona una nueva y prometedora aproximación al campo de la Ingeniería Civil, donde es extremadamente importante medir con precisión diferentes procesos.

Sin embargo, la Visión Artificial es un campo muy amplio que abarca multitud de técnicas y objetivos, y definir una aproximación de desarrollo sistemática es problemático.

En esta tesis se propone una nueva metodología para desarrollar estos sistemas considerando las características y requisitos de la Ingeniería Civil.

Siguiendo esta metodología se han desarrollado dos sistemas:

Un sistema para la medición de desplazamientos y deformaciones en imágenes de ensayos de resistencia de materiales. Solucionando las limitaciones de los actuales sensores físicos que interfieren con el ensayo y solo proporcionan mediciones en un punto y una dirección determinada.

Un sistema para la medición de la trayectoria de peces en escalas de hendidura vertical, con el que se pretende solucionar las carencias en el diseño de escalas obteniendo información sobre el comportamiento de los peces.

Estas aplicaciones representan contribuciones significativas en el área, y demuestran que la metodología definida e implementada proporciona un marco de trabajo sistemático y confiable para el desarrollo de sistemas de Visión Artificial en Ingeniería Civil.

RESUMO

A Visión Artificial proporciona unha nova e prometedora aproximación ó campo da Enxeñería Civil, onde é extremadamente importante medir con precisión diferentes procesos.

Sen embargo, a Visión Artificial é un campo moi amplo que abarca multitude de técnicas e obxectivos, e definir unha aproximación de desenvolvemento sistemática é problemático.

En esta tese proponse unha nova metodoloxía para desenvolver estes sistemas considerando as características e requisitos da Enxeñería Civil.

Seguindo esta metodoloxía desenvolvéronse dous sistemas:

Un sistema para a medición de desprazamentos e deformacións en imaxes de ensaios de resistencia de materiais. Solucionando as limitacións dos actuais sensores físicos que interfiren co ensaio e só proporcionan medicións nun punto e nunha dirección determinada.

Un sistema para a medición da traxectoria de peixes en escalas de fenda vertical, co que se pretende solucionar as carencias no deseño de escalas obtendo información sobre o comportamento dos peixes.

Estas aplicacións representan contribucións significativas na área, e demostran que a metodoloxía definida e implementada proporciona un marco de traballo sistemático e confiable para o desenvolvemento de sistemas de Visión Artificial en Enxeñería Civil. .

INDEX

INDEX OF FIGURES

INDEX OF TABLES

Every problem has a gift for you in its hands

Richard Bach

I. Introduction

1 Motivation and Objectives

Nowadays, intelligent systems stand for the best approach to handle several problems involving optimization, complexity, imprecision or uncertainty in noisy and non-stationary environments.

As these new paradigms evolve, the way the new technologies interact with the real world is also changing quickly, and the perception of the world through vision seems to be the key to a new type of computation and robotics, where machines would be able to deal for the first time with data and representations feasible by humans.

Using this shared level of perception, everyone and everyday applications based on images are emerging in a world where personal photographic devices are a new social standard.

Computer Vision (*CV*), the field focused on extracting knowledge from images with computers, will be located in the center of this technological revolution, given that imagery will be an essential type of information in the future (and it is already so).

Therefore, large image and video databases will be common and they will serve as key sources of information for people in their everyday life, as well as for industry or professionals' applications. To operate with all this information, new and sophisticated imaging and *CV* techniques will be required and are already being developed.

In Civil Engineering, and other practical areas such as biomedicine, where it is extremely important to measure with precision real processes, the evolution of *CV* systems is establishing a new paradigm providing new solutions to solve traditional problems, achieving more accurate results at lesser cost, opening new possibilities and making it possible to obtain accurate information from complex and non-deterministic processes.

Table 1: Benefits and drawbacks of *CV* systems [1].

Advantages	Disadvantages
- Generation of precise descriptive data	- Object identification being considerably more difficult in unstructured scenes.
- Quick and Objective	
- Reducing tedious human involvement	- Artificial lightning needed for dim or dark conditions
- Consistent, efficient and cost effective	
- Automating many labour intensive process	
- Easy and quick, consistent	
- Non-destructive and undisturbing	
- Robust and competitively priced sensing technique	

CV techniques have been widely used in industrial and medical scenarios, where they can offer several benefits as summarized in Table 1.

As it will be discussed later, a visual image is inherently ambiguous and perception is essentially a matter of resolving ambiguities by using knowledge.

Therefore, vision can be seen as a process that produces from images of the external world a description that is useful to the viewer and that is not cluttered with irrelevant information.

From the point of view of an applied information system, the purpose of vision itself cannot be reduced to the description of scenes which will be used by high level processes. It must be considered as an act toward the external world in order to achieve a particular goal. The purpose of vision is to solve a particular visual task using appropriate information, representations, algorithms and hardware.

In such domain, the goal of the system is well defined and the job of the designer is to find all the components he needs in order to achieve his goal.

To develop a *CV* system in an efficient way, the designer has to be a vision engineer more than a vision expert. Therefore, it is necessary to have a methodology which allows him to evaluate all the components of the application in such a way that he can start with the global goal and progressively refine the architecture of the system as well as its components.

To study *CV* systems through a systematic approach, the first step is to describe a general *CV* system, understood as a set of general techniques using a general theory. Then, it is necessary to identify the main tasks, subtasks, processes, components and involved relations in order to define an engineering, systematic and repeatable approach.

There have been different attempts to establish a general theory for vision, such as the Marr paradigm [2], which states that the principles of vision can be studied by

considering describing scenes as the final purpose of vision. Using this goal, Marr formulates three different levels of understanding:

- Computational theory (What is the goal of the computation?).
- Representation and algorithm (How can they be implemented?).
- Hardware implementation (How can it be done physically?).

This theory allows the decomposition of the higher level explanation of the studied phenomenon. However, considerations about context-dependence can and should arise at more than one level of analysis and may have different answers at different levels.

Indeed, some authors had stated that there is not an unique computational theory of vision [3]. According to these authors, for each architecture choice of a *CV* system, it is possible to develop a computational theory. Furthermore, it can be said that there is no methodology which helps the designer in choosing the architecture of the system and there is no reason to suppose the existence of a common representation of such a high level as the 3D world model.

Currently, it should be accepted that no general vision algorithm exists and a *CV* system must be developed using a case-by-case approach, in an *ad hoc* way [3].

Using a systematic approach in Computer Vision is very problematic but also necessary, since it is obvious that the development cost would be too high in the event that exploring all possible techniques to find the optimal solution was required.

On the other hand, performing a design based on empiric knowledge is also an undesired solution since the process would be based mostly on the designers experience instead of an engineering process.

The objective of this thesis is to define a methodology to develop a *CV* system in Civil Engineering attending to the specific characteristics of the field.

This will be done with no need of a general theory or a unique common principle, by studying in a practical point of view the general problems, elements, and techniques of *CV* fields; defining the relations and the general processes involved when analysing Computer Vision from a Civil Engineering perspective.

This analysis will be performed using a similar perspective than for inspection systems from industry; starting from the problem domain and using empiricism to identify and decompose a task into subtasks; detailing then the main elements and actions and taking into account the properties of the objects and the requirements of the solution.

To this end, *CV* systems will be analysed as entities composed of interdependent elements such as the observer, the observed scene, the acquisition and the processing elements which must be understood in terms of their relationships inside the global entity.

The result of this analysis is not only a methodology providing the cycle of actions required to design a *CV* system, but also a characterization of Computer Vision and a practical guide in the task of designing a *CV* system for Civil Engineering.

2 Thesis Overview

This thesis has been organized in different chapters as follows:

- In *Chapter II* the fundaments of Computer Vision are introduced, defining its relations with other fields and then examining the main topics and techniques.

- In *Chapter III* the proposed methodology is introduced. First, *CV* systems are analysed from the perspective of Civil Engineering; defining their common elements and processing stages. Then, the knowledge which may be introduced in the system will be analysed; exploring the assumptions and camera models which may lead to obtain robust and real measurements from images, together with other types of knowledge required, such as information about the objects or classes of interest and the knowledge extracted by the system. Then, a general methodology to design the system will be analysed. Finally, the main aspects of the experiment design will be discussed.
- *Chapter IV* shows one of the results obtained with this thesis: a system developed with the proposed methodology to analyse strength tests based on Block-Matching techniques.
- In *Chapter V* another system developed with the proposed methodology is presented. This is a *CV* system to measure the trajectory of fishes in fishways based on different image processing techniques.
- In *Chapter VI* the final conclusions and the future lines of work within the aim of this thesis are studied.

3 Contributions

Several parts of the work in this thesis, or inspired by it, were published in different media. The main contributions are listed below:

International Journals with Impact Index

A. Rodriguez, M. Bermudez, J. Rabuñal, J. Puertas, J. Dorado, and L. Balairon, "Optical Fish Trajectory Measurement in Fishways through Computer Vision and Artificial Neural Networks," Journal of Computing in Civil Engineering, vol. 25, pp. 291-301, 2011. JCR in 2011: 1.337. Ranking 23/118(Q1) in Engineering, civil; 49/99(Q2) in Computer Science, Interdisciplinary Applications.

J. Puertas, L. Cea, M. Bermudez, L. Pena, A. Rodriguez, J. Rabuñal, et al., "Computer application for the analysis and design of vertical slot fishways in accordance with the requirements of the target species," Ecological Engineering, vol. 48, pp. 51-60, 2012. JCR in 2012: 2.958. Ranking 38/136(Q2) in Ecology; 10/42(Q1) Engineering, Environmental; 42/210(Q1) in Environmental Sciences.

A. Rodriguez, J. Rabuñal, J. L. Perez, and F. Martinez-Abella, "Optical Analysis of Strength Tests Based on Block-Matching Techniques," Computer-Aided Civil and Infrastructure Engineering, vol. 27, pp. 573–593, 2012. JCR in 2012: 4.460. Ranking 2/100(Q1) in Computer Science, Interdisciplinary Applications; 1/57(Q1) in Construction & Building Technology; 1/122(Q1) in Engineering, Civil; 1/30(Q1) in Transportation Science & Technology.

A. Rodriguez, C. Fernández-Lozano, J. Dorado, and J. Rabuñal, "2-D Gel Electrophoresis Image registration using Block-Matching Techniques and Deformation Models," Analytical Biochemistry, 2014 (Accepted, Publication Pending). JCR in 2012: 2.582. Ranking 32/75(Q2) in Biochemical Research Methods; 166/290(Q3) in Biochemistry & Molecular Biology; 26/75(Q2) in Chemistry, Analytical.

International Journals without Impact Index

A. Freire, J. A. Seoane, A. Rodriguez, C. Ruiz Romero, G. López-Camos, and J. Dorado, "A Block Matching based technique for the analysis of 2D gel images," Studies in health technology and informatics, vol. 160, pp. 1282-1286, 2010.

A. Rodríguez , C. Fernández-Lozano, and J. Rabuñal, "Spot matching in 2D electrophoresis experiments," International journal of Data Analysis and Strategies, vol. 5, pp. 198-213, 2013.

International Conferences

A. Rodríguez, J. Rabuñal, J. Dorado, and J. Puertas, "Overflow Channel Study using Image Procesing," presented at the International conference on Image Processing & Communications (IPC), Bydgoszcz, Poland, 2009.

A. Rodriguez, C. Fernandez-Lozano, J.-A. Seoane, J. Rabuñal, and J. Dorado, "Motion Estimation in Real Deformation Processes Based on Block-Matching Techniques," presented at the IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), 2011.

A. Rodriguez, C. Fernandez-Lozano, J.-A. Seoane, J. Rabuñal, and J. Dorado, "Analysis of Deformation Processes Using Block-Matching Techniques," presented at the International Conference on Computer Vision Theory and Applications, Roma, Italia, 2012.

A. Rodriguez, J. Rabuñal, M. Bermudez, and A. Pazos, "Automatic Fish Segmentation on Vertical Slot Fishways Using SOM Neural Networks," presented at the International work Conference on Artificial Neural Networks (IWANN), Tenerife, España, 2013.

Publications Being Processed

A. Rodriguez, M. Bermúdez, J. Rabuñal, and J. Puertas, "Fish tracking in vertical slot fishways using computer vision techniques," 2014.

A. Rodriguez and J. Rabuñal, "Measuring Deformations in Biomedicine with Block-Matching Techniques: A Survey," 2014.

It is not certain that everything is uncertain

Blaise Pascal

II. Fundaments

1 The Image

An image can be seen as a signal subset, a signal being a function that conveys information about the behavior of a physical system or attributes of some phenomena.

Although signals can be represented in many ways, in all cases the information is contained in a pattern of variations of some form, and that information is typically transmitted and received over a medium.

An image, as we understand, is a spatially varying pattern representing the light reflected by a surface or surfaces in a scene, captured and usually digitalized in a physical device like a camera [4].

Images represent the light reflected by a scene and projected to a 2D plane called the image plane. Then, the incident light is codified in a discrete range and in a

discrete set of positions or pixels, so the values for each pixel represent the intensity or color of light received in a point of the camera.

Therefore, an image can be defined as the distribution of the refraction indexes of the objects in a scene.

However, the concept of image can be extended to a frequency representation, instead of a spatial one, in which each image value at an image position F represents the amount that the intensity values in the image vary over a specific distance related to F. [5].

Although the image concept is usually referred to a representation of the visible part of the light, it can also be extended to different bands of the electromagnetic spectrum such as the infrared light, and to different physical properties, such as the response of objects to sound.

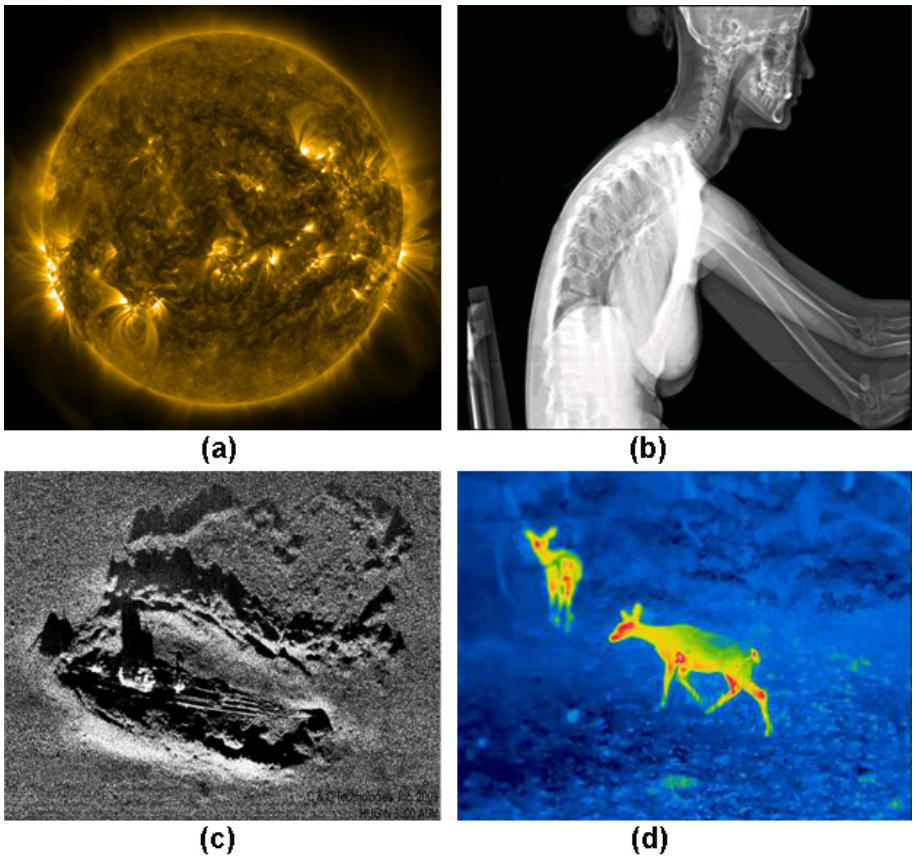

Figure 1: Images.

(a) Magnetic image from the Sun taken by the Solar Dynamics Observatory [6]. (b) Radiography of the column [7]. (c) SONAR image of a German U-Boat U-166 collected by the HUGIN 3000 [8]. (d) Infrared imaging for deer hunting [9].

From a physical and general point of view, an image is defined as a function $C(x, y, t, \lambda)$, representing the energy distribution of an image source of radiant energy at spatial coordinates (x, y), at time t and with a wavelength λ. [10].

Finally, mathematically speaking, images may be defined as a two-dimensional function $f(x, y)$, where x and y are spatial (plane) coordinates, and the amplitude of

f at any pair of coordinates *(x, y)* represent a property such as the gray level (intensity of light) at that point. [11].

Therefore, an image can be represented as a matrix, a function or a statistical distribution. It can be analysed and processed using several mathematical and statistical tools. Additionally, understanding the mathematical processes of image formation, information about the real world can be obtained from images.

Nowadays, applications using images and obtaining information from images are becoming more and more important and are already present in a significant number of everyday applications.

The field focused on analysing and extracting information from images is called Computer Vision (*CV*).

2 Computer Vision

As humans, we perceive the three-dimensional structure of the world around us with apparent ease. Humans can tell the shape and translucency of each object in a scene through the subtle patterns of light and shading that play across its surface and effortlessly segment each of these objects from the background, recognize the class of the object and register it with the abstract concept associated in our brain with this class.

Human sight can also perceive the position and size of the object in the real world, and even track a moving object; interpolating its trajectory when is hidden from our sight or predict its future position.

Vision is the most advanced of our senses and more than a 70% of the information that we receive from the world is acquired from vision [12].

Scientifics have spent decades trying to understand how the visual system works and, even though they can devise optical illusions to tease apart some of its principles (Figure 2), a complete solution to this puzzle remains elusive [13].

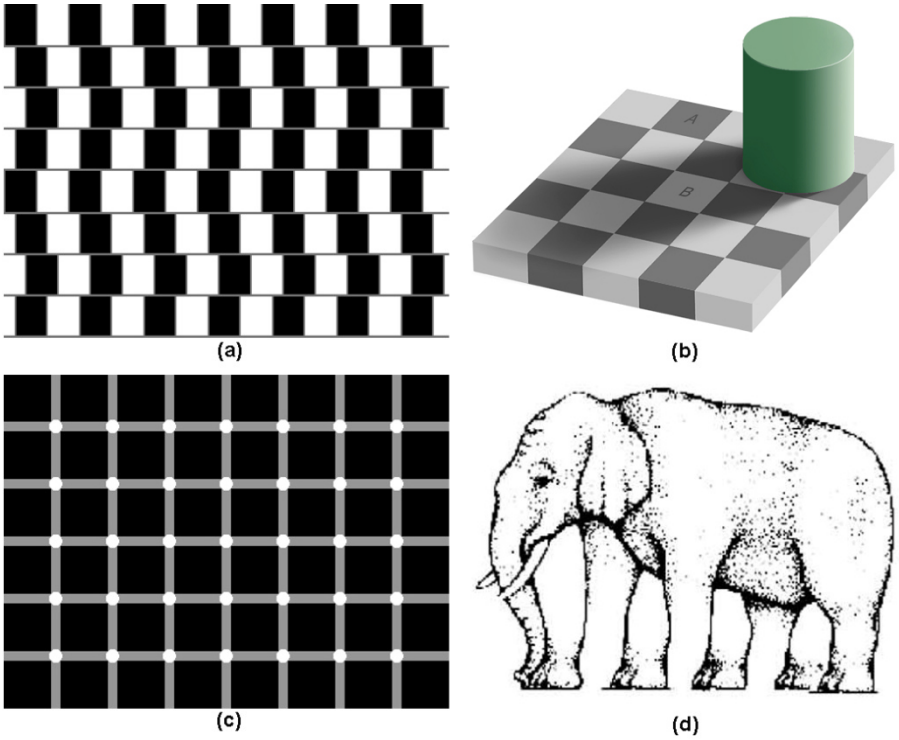

Figure 2: Optical illusions.

(a) Kindergarten illusion. The lines parallels and horizontals [14]. (b) Checker shadow illusion. The squares A and B have the same color [15]. (c) The scintillating grid illusion. The points in the intersection seem to change its color [16]. (d) Elephant illusion. How many legs has the elephant? [17].

Visual information is currently also an important source of information for computer systems. It is providing a new way of thinking, where images are used not only to extract information from the world but also to perform actions in it.

Computer Vision is a field of computer science born in 1970s that includes methods for acquiring, processing, analysing, and understanding images and, in general, high-dimensional data from the real world in order to produce numerical or symbolic information, e.g., in the form of decisions.

Figure 3 shows a timeline with some of the main contributions in Computer Vision.

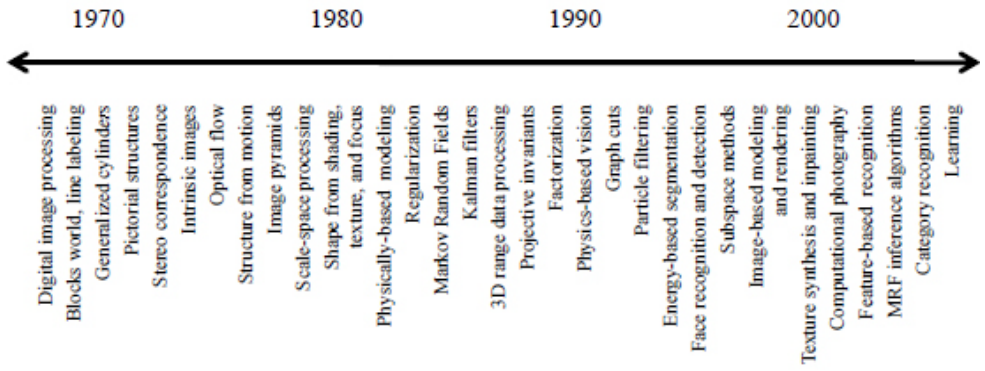

Figure 3: Computer Vision timeline [18].

Computer Vision typically requires a combination of low level image processing to enhance or modify the image and higher level pattern recognition or image understanding techniques.

As a scientific discipline, Computer Vision is concerned with the theory behind artificial systems that extract information from images. As a technological discipline, on the other hand, Computer Vision seeks to apply its theories and models to the construction of *CV* systems [19-21].

Unlike humans, who are limited to the visual band of the electromagnetic (*EM*) spectrum, imaging machines cover almost the entire *EM* spectrum, ranging from gamma to radio waves. They can operate on images generated by sources that

humans are not accustomed to associating with images. These include ultrasound, electron microscopy and computer-generated images [11].

Nowadays, *CV* systems perform with success tasks such as computing a 3D model of an environment from partially overlapping photographs. It is also possible, given a large enough set of views of a particular object, to create a dense 3D surface model of the object. Using *CV* systems, we can track a person moving against a complex background. We can even, with moderate success, attempt to find and name all of the people in a photograph using face detection and recognition [13].

Examples of these *CV* tasks are shown in Figure 4.

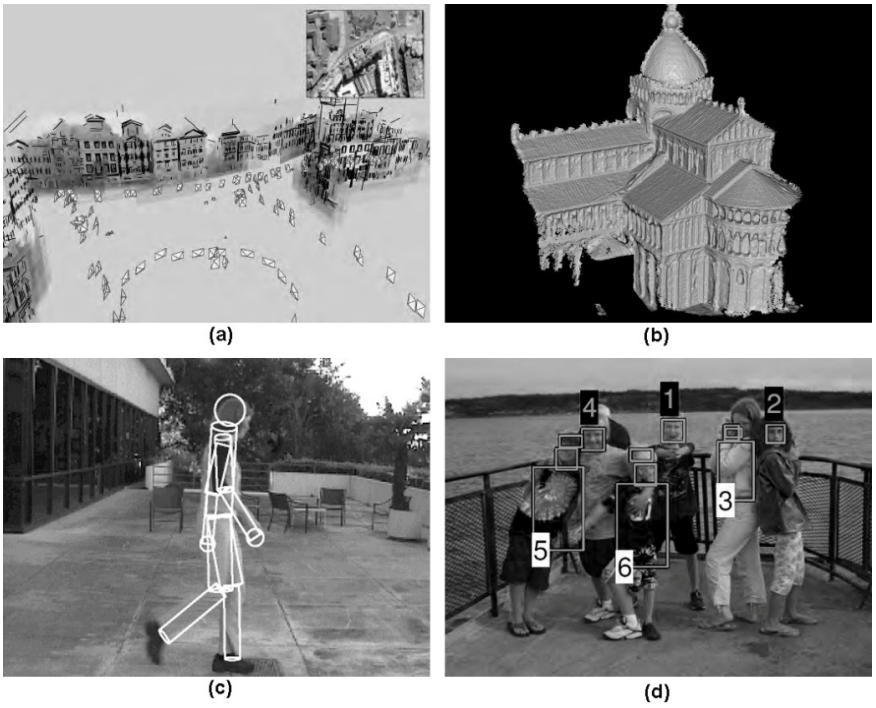

Figure 4: Some examples of *CV* tasks.

(a) Computing a 3D model of an environment from partially overlapping photographs. (b) Creating a dense 3D surface model from a large enough set of views of the particular object. (c) Person tracking against a complex background. (d) Naming all of the people in a photograph [18].

However, despite all of these advances, a computer interpreting an image at the same level as a two-year-old child (capable for example of counting all of the animals in a picture) is beyond the state of the art nowadays [13].

But, why *CV* algorithms are so error prone when humans and animals do this so effortlessly?

The answer to this question is a key issue on Computer Vision; the problem is related to the image data and to the fact that Computer Vision is a backward process which tries to describe the world from one or more images and to reconstruct its properties, such as shape, illumination, and color distributions.

Computer Vision can be therefore understood like the opposite to the forward models developed in physics and in computer graphics. These fields model how objects move and animate; how light reflects off their surfaces, is scattered by the atmosphere, refracted through camera lenses, and finally projected onto image plane. These models are much evolved and they can currently provide an almost perfect illusion of reality [13].

However, in Computer Vision, and while some pixel-level processing techniques and some tasks can be performed more accurately by computers than by natural brains, the actual problem may lie in the image understanding processes.

As mentioned in section 1, an image represents the distribution of light returned from one or more objects and received by a sensor (which depends on the light source intensity and on the relative positions of the light source, the objects and the observer). This implies an inherent presence of noise in the process and, as the information proceeds from a 3D to 2D projection of the light, we only receive a part of the information.

The problem lies in the fact that there are infinite possible scenes corresponding to the image that we receive, so it is necessary an interpretation of the image to deal with the correct answer.

Finally, there are also other problems such as the huge amount of data to be processed, the limitations of the optical systems, etc.

Thus, while a human or an animal brain has capacity for applying learned models and uses generalization or inference processes, machines cannot currently obtain a meaningful representation of what they are seeing at the same level as we do.

Currently, there are in Computer Vision several applications to solve different problems, such as detecting an object of interest, tracking and moving objects or performing different real world measurements from a sequence recorded with a calibrated camera.

However, it is the ability of integrating all this processing in a general and meaningful framework what is beyond the current state of the art.

As an example, in the task of avoiding an object in a collision course (e.g. a ball in a game) humans probably apply a segmentation process using color and background information together with motion segmentation and distance segmentation. The human brain discards the non-relevant information, recognizes the ball, track its motion and predicts its future trajectory; using this information to react in an active way to modify the subject position, avoiding the collision.

A robot can also be programmed to solve a similar task with moderate success; using a stereo vision system together with segmentation and pattern recognition techniques to detect the ball. Then, a probabilistic approach can be used to predict

the movement of the ball and react to it by using knowledge about its own position.

However, what makes the difference between natural vision and Computer Vision is the ability of putting together the previous knowledge with automatic processes of the data in a complex and comprehensive way.

Humans are able, therefore, to react dynamically to changes in the environment; prioritizing the received information, modifying the perception process itself and selecting the focus of attention. We can use a dynamical brain model of the background or associate image features, real objects and mind concepts. Furthermore, we can decide a course of action which may involve a totally different cognitive process depending on the context, putting together the previous knowledge with automatic processes of the data.

2.1 Related Fields

Computer Vision is a very interdisciplinary field. There are currently several related fields with Computer Vision, and usually there is no general agreement among authors regarding where one of these areas stops and other starts, or about what relationship exists between two of these areas [11].

An example of diagram with the most important of these fields is shown in Figure 5.

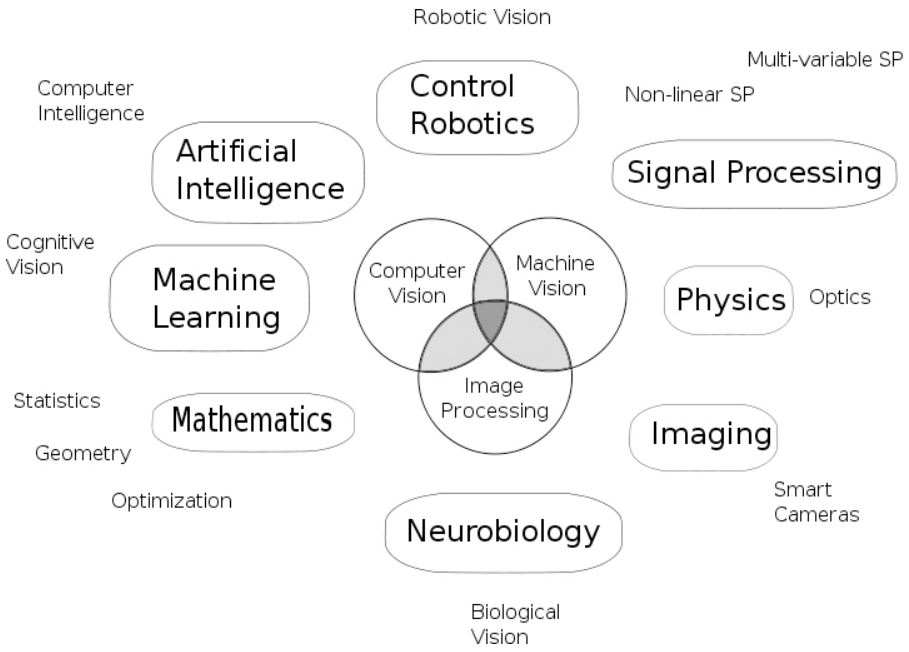

Figure 5: Computer Vision related fields [22].

As these fields are usually presented, there is a significant overlap in the range of techniques and applications between the Computer Vision, machine vision and image processing areas. Furthermore, the basic techniques that are used and developed in these fields are sometimes almost identical and the differences may be only relevant for marketing reasons.

The following characterizations appear relevant, but should not be taken as universally accepted:

- **Image Processing:** It is focused on 2D images, studying how to transform one image into another, e.g., by pixel-wise operations such as contrast enhancement, local operations such as edge extraction, noise removal, or geometrical transformations such as rotating the image. This characterization implies that image processing neither require assumptions nor produce interpretations about the image content. Usually, the distinction with

Computer Vision is made by defining image processing as a discipline in which both the input and output of the process are images [13].

- **Machine Vision:** It is the process of applying a range of technologies and methods to provide imaging-based automatic inspection, process control and robot guidance in industrial applications [23, 24]. From a functional point of view, it can be seen as the analysis of images to extract data for controlling a process or activity [25].

In this thesis we will use the Computer Vision term to refer in a wide sense to any computer system used to extract any kind of knowledge from images, using image processing and higher level techniques, and it would include totally the machine vision field.

Due to its objectives (understanding or perceiving the environment) and some of the used techniques (Artificial Neural Networks, learning techniques), Computer Vision is usually seen as a part of the Artificial Intelligence (*AI*) field but, again, this is not universally accepted and *CV* applications and investigations appear usually in specific journals and not in journals focused on Artificial Intelligence.

Important related fields with Computer Vision are: Physics, which explains the behavior of optics and the process by which light interacts with surfaces. Imaging, which is primarily focused on the process of producing images. Neurobiology, which studies the biological vision system. Robotics, which can use visual information of the environment to navigate or react in the environment. And many others.

There is also a special relationship between Computer Vision and signal processing. This field is primarily focused on processing one-variable temporal signals, but it can be extended to process images.

Some of the signal processing techniques were developed within Computer Vision, in a subfield of signal processing which can be considered as a part of Computer Vision, and which can be identified with image processing.

There is also a branch of signal processing, the pattern recognition, which uses various methods to extract information from signals using statistical approaches. A significant part of this field is devoted to applying these methods to image data and one of the typical issues of the Computer Vision, the classification problem, is usually studied from the point of view of pattern recognition.

Besides the above mentioned views on Computer Vision, many of the related research topics can also be studied from a purely mathematical point of view. For example, many methods in Computer Vision are based on statistics, optimization or geometry.

Finally, a significant part of the field is devoted to the implementation aspect of Computer Vision; how existing methods can be realized in various combinations of software and hardware, or how these methods can be modified in order to gain processing speed without losing too much performance.

2.2 Main Topics and Techniques

Since Computer Vision is such a broad field, involving going from images to a structural description of the scene and including different theoretical areas and several application fields, it is extremely difficult to classify the different topics of the area and establish the relations between them.

Figure 6 represents a possible taxonomy of the main topics in Computer Vision. The different topics are disposed horizontally in terms of which major component they address, in addition to vertically according to their dependence.

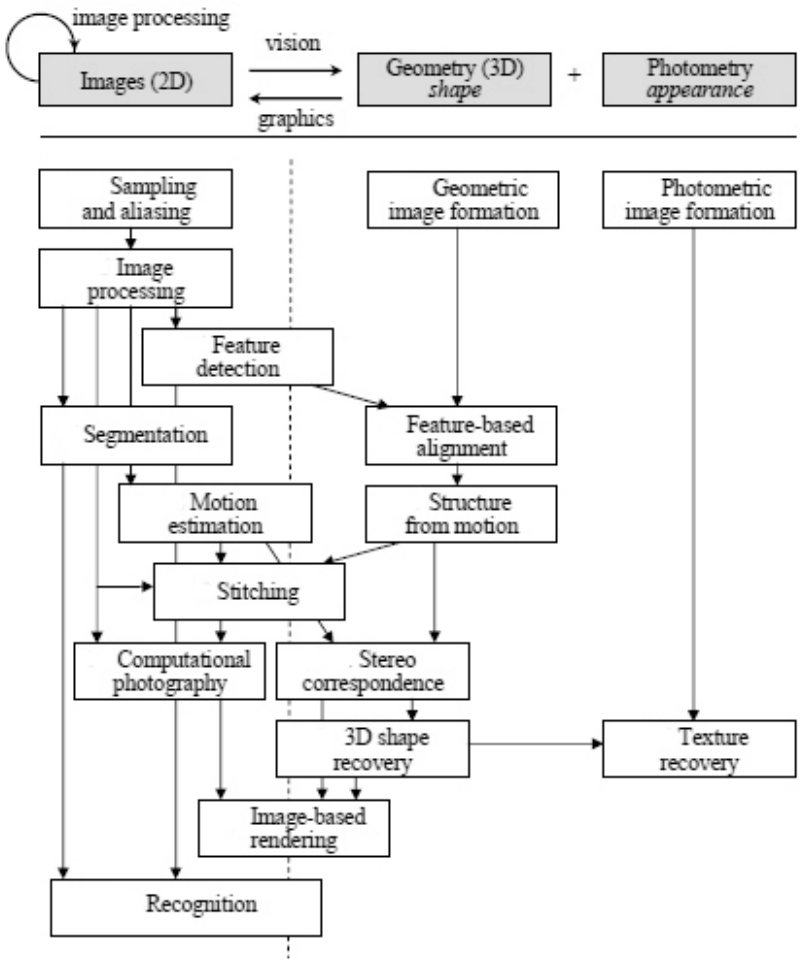

Figure 6: Taxonomy of main *CV* topics.

Topics are roughly positioned along the left–right axis depending on whether they are more closely related to image-based (left), geometry-based (middle) or appearance-based (right) representations, and on the vertical axis by increasing level of abstraction. The whole figure should be taken with a large grain of salt, as there are many additional subtle connections between topics not illustrated here [18].

Next, each of these topics will be explained together with the most common techniques used in it. This scheme should also be taken only as a reference, since many of the named techniques can be used in other topics or with different purposes.

2.2.1 Geometric Image Formation

This topic is mainly focused in geometrical structures (such as points, lines and planes), and also in the projection and geometrical transformations of these features (such as affine or projective transformations). This includes the theory behind world to image projection in the image formation, used in calibration and distortion removal; and it is in the background of multi-view reconstruction and shape recovery applications. [26, 27].

2.2.2 Photometric Image Formation

This topic is related with radiometry and optics. It is mainly focused on the relation of image values with lightning. This includes the study of illumination sources, surface properties (using models of reflectance and shading) and camera optics. [28, 29].

2.2.3 Sensor Image Formation

This topic is mainly focused on the study of image sensor properties, including sampling, aliasing, color sensing and in-camera compression [30].

Figure 7 shows a simple version of the processing stages that occur in modern digital cameras.

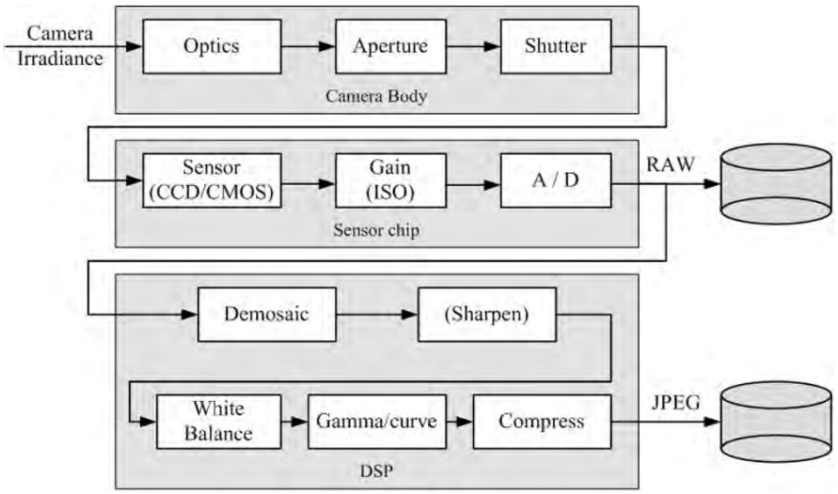

Figure 7: Processing stages in a digital camera [18].

2.2.4 Image Processing

As explained in section 2.1, image processing deals mainly with operators that map pixel values from one image into another. These techniques are usually included as a part of the signal processing area.

Interest in digital image processing methods stems from two principal application areas: improvement of pictorial information for human interpretation; and processing image data for storage, transmission, and representation for autonomous machine perception [11].

Examples of image processing operations include exposure correction, color balancing, reduction of image noise, increasing of sharpness, and others.

Most *CV* applications require care in designing the image processing stages in order to achieve acceptable results.

Point Operators

The simplest kinds of image processing transforms are point operators, where each output pixel value depends on only the corresponding input pixel value (plus, potentially, some globally collected information or parameters).

Point operators include brightness and contrast adjustments, color correction and transformations as well as many other techniques.

An example of a simple point operator is the gamma correction, where a constant value is added to each pixel to increase the luminosity in the image. A more complex operation is the histogram equalization, a technique to improve the appearance of an image by redistributing the intensity values to cover the full intensity spectrum in order to increase the contrast [31].

Neighborhood Techniques

Neighborhood operators are similar to point operators but, in this case, new pixel values depend on a small number of neighboring input values.

Some of these techniques include:

- **Linear Filtering:** They are the most used operators in Computer Vision. They consist on an operation where the output value for each pixel is determined as a weighted sum of input pixel values in the neighboring of the original pixel. The entries in the weight kernel or mask are often called the filter coefficients. These operators can be used to filter images, sharpen details, accentuate edges, or remove noise. [13].
- **Morphological Image Processing:** Also called mathematical morphology, it can be seen as a tool for extracting image components that are useful in the

representation and description of region shape; such as boundaries, skeletons and the convex hull [11].

Mathematical morphology was originally developed for binary images and was later extended to grayscale images. In morphological techniques, an image is seen as a subset of an Euclidean space and morphological operations are based on set theory. In these operations, images are transformed using a simple pre-defined shape called structuring element, which can be considered as an image itself.

Therefore, the image is operated with the structured element in a similar way as a kernel in linear filtering; being the image modified according to the correspondence of the patterns in the image with the chosen pattern in the structuring element.

Fourier Transforms

The Fourier Transform (*FT*) is an important image processing tool which is used to decompose an image into its sine and cosine components. The output of the transformation represents the image in the Fourier or frequency domain, while the input image is the spatial domain equivalent [5]. The FT is used in a wide range of applications; such as image analysis, image filtering, image reconstruction and image compression.

The Fourier Transform is used if it is necessary to access the geometric characteristics of a spatial domain image. Because the image in the Fourier domain is decomposed into its sinusoidal components, it is easy to examine or process certain frequencies of the image, thus influencing the geometric structure in the spatial domain.

- **Fast Fourier Transform (*FFT*):** It is one of the most used techniques to perform a Fourier Transform. One of its advantages is that it can be used to perform efficiently large-kernel convolutions in a time that is independent of the kernel's size [6].

Multiresolution Techniques

Multiresolution techniques are based on psychophysics and physiological experiments which have shown that multiscale transforms seem to appear in the visual cortex of mammals [2]. The main idea is to use representations of the image with scales to analyse the images.

- **Pyramids:** Also called pyramidal decompositions, pyramids are used to process an image using a hierarchy of scales. This technique consists in resampling iteratively the image, so it can be analysed using a coarse to fine approach. Pyramids are useful to avoid local minimums and to reduce the computational time.

  The best known and most widely used pyramid in Computer Vision is the one published by Burt and Adelson [32].
- **Wavelets:** The wavelets transform is a spatial and frequency representation of an image. It was introduced in Computer Vision by Mallat [33]. They are calculated convoluting the image with a set of functions generated from a main wavelet (a function which integrates to zero and can be typically visualized as a brief oscillation) and a scaling function. One of the most important uses of wavelets is in data compression. Wavelets can be calculated using a Fast Fourier Transform (*FFT*).

Geometrical Transforms

Geometrical transformations assign to each pixel a new position depending on its initial position and a transformation function.

As mentioned before, they are also studied in geometric imaging. They are considered as a part of physics, maths and projective geometry.

Some examples of geometric transformations are translations, rotations or projections [26, 34].

Optimization Techniques

- **Regularization Methods:** They are based on the construction of a global energy function that describes the desired characteristics of the solution. Then a minimum energy solution should be founded using sparse linear systems or related techniques.
- **Bayesian Statistics:** They are based on modeling the noisy measurement process that produced the input images as well as prior assumptions about the solution space, which are often encoded using a Markov Random Field (*MRF*, a stochastic process has the Markov property if the conditional probability distribution of future states of the process depends only upon the present state).

2.2.5 Feature Detection

Feature detection and matching is an essential component of many *CV* applications. Point features can be used to find a sparse set of corresponding

locations in different images. These techniques are often used as a pre-cursor to computing camera pose, aligning images or object recognition.

A typical feature based analysis can be separated in three stages [13]:

1. Feature detection.
2. Feature description. After detecting features, usually, a feature description stage is involved where each feature is transformed in a more compact and stable (invariant) descriptor which can be matched to other descriptors.
3. Find the correspondence of the features, there are two main approaches to solve this problem:
    a. Find features in one image and track them in other images using local search techniques [35, 36].
    b. Detect features in all the images under consideration and then match them using their local appearance [22].

More information about finding correspondences in images can be found in section 2.2.7 and section 2.2.9.

Complex Features: Regions Edges and Lines

While interest points are useful for finding image locations that can be accurately matched in 2D; regions, corner or edge points are far more plentiful and often carry important semantic associations.

While finding regions of interest is itself a wide field of Computer Vision (see section 2.2.6), most edge detection techniques are fast and easy to implement.

Qualitatively, edges occur at boundaries between regions of different color, intensity or texture. Under such conditions, a reasonable approach is to define an edge as a location of rapid variation of these features.

Detecting edges is usually performed by using intensity gradients that are image derivatives. The theory behind this is that a high value in the derivative will mark a fast change and therefore a possible edge between two image regions.

- **Linear Filters:** Because differentiation is a linear operation, linear filters are commonly used to detect edges [7]. The Gaussian is the main separable circularly symmetric filter and it is used in most edge detection algorithms. Other possibilities are directional filters such as the Canny, Sobel, Prewitt, Roberts cross and Frei-Chen operators. Also second order differentiation is commonly used, applying operators such as a Laplacian of Gaussian (*LoG*) kernel and then analysing the resulting zero-crossings.
- **Hough Transform:** It is a feature extraction technique, whose purpose is to find imperfect instances of objects within a certain class of shape by a voting procedure. Although Hough transform has been extended to identifying positions of arbitrary shapes, it tends to be most successfully applied to line finding in sets of edge points [37]. It works taking each candidate point and determining all lines that could pass through that token. The lines present in the image will be carried out because they pass through many points and therefore they will have a high score.

2.2.6 Segmentation

Image segmentation is the task of finding groups of pixels that correspond to the same class, usually the same real object.

In statistics, this problem is known as cluster analysis and is widely studied. In Computer Vision, image segmentation is one of the oldest and most widely studied problems and there are hundreds of different approaches to solve it.

The simplest possible technique for segmenting an image is to select a threshold and then compute connected components. Unfortunately, a single threshold is rarely sufficient to segment the whole image because of lighting and intra-object statistical variations.

Techniques such as edge detection, point operations to threshold the image or optimization techniques can be used in image segmentation. Also techniques from Artificial Intelligence such as Artificial Neural Networks *(ANN)* [38] or genetic algorithms can deal with the problem.

There are, in addition, different specific techniques which can be used. They can be classified as follows [13]:

Background Subtraction Techniques

One of the first steps in many applications is to remove the background from the image to reduce the undesired information and to make easier to perform further operations. This is usually performed by modeling the background using *a priori* knowledge and is very useful in experiments in a controlled environment. Background subtraction techniques can be seen as an inverse segmentation process, where we try to find the objects of interest by subtracting the non-interesting objects from the image [39].

Active Contours

This technique is focused on finding curves corresponding to object boundaries using dynamic structures which will evolve to cover a part of the image corresponding to a region of interest.

Active contours allow a user to roughly specify a boundary of interest (or just an initial point). Then, the system will evolve the contour towards a more accurate location and it will track it over time.

- **Snakes:** They are energy-minimizing two-dimensional spline curves that evolve (move) towards image features such as strong edges. They can be studied as a special type of optimization technique [9].
- **Scissors:** They are techniques to optimize the contour of an object of interest in real time as the user is drawing. To compute the optimal curve path, low costs are associated with edges that are likely to be boundary elements. Next, as the user traces a rough curve, the system computes the lowest cost path between the starting point and the current mouse location [40].

Region Growing

Region growing techniques are based on using dynamic areas which may evolve to represent a region of interest. Region growing techniques are similar to active contours, but they are not focused on boundaries.

- **Split and Merge:** Split and merge techniques proceed either by recursively splitting the whole image into pieces based on region statistics or, inversely, merging pixels and regions together in a hierarchical fashion [41].

It is also possible to combine both splitting and merging by starting with a medium-grain segmentation; and then allowing both merging and splitting operations

- **Watershed:** This technique segments an image into several catchment basins, which are the regions of an image (interpreted as a height field or landscape) where rain would flow into the same lake. An efficient way to compute such regions is to start flooding the landscape at all of the local minima and to label ridges wherever differently evolving components meet [42].

Mean Shift and Mode Finding Techniques

Mean-shift and mode finding techniques model the feature vectors associated with each pixel (e.g., color and position) as samples from an unknown probability density function and then try to find clusters (modes) in this distribution.

- **K-means:** This technique implicitly models the probability density as a superposition of spherically symmetric distributions. It does not require any probabilistic reasoning or modeling. The algorithm is given the k number of clusters it is supposed to find and then it iteratively updates the cluster center locations based on the samples that are closest to each center [43].
- **Mixtures of Gaussians:** This technique is also based on a given number of clusters, where for each cluster center a covariance matrix, whose values are re-estimated from the corresponding samples, is used. This latter approach corresponds to iteratively re-estimating the parameters for a mixture of Gaussians density function [43].

Cuts and Energy Based Techniques

These techniques are based on finding a partition of the image (cut) by minimizing some metric or energy based term.

- **Normalized Cuts:** This technique examines the similarities between nearby pixels and tries to separate groups that are connected by weak affinities. This technique minimizes a normalized measurement of region similarity formed by the sum of weights of its pixels. Then it uses this minimum cut as a segmentation criterion. [44].
- **Graph Cuts:** These techniques are based on processing an image as a graph, where basically each pixel in the image is viewed as a node in a graph. Here, edges are connections between nodes with weights corresponding to how alike two pixels are, given some measure of similarity, as well as the distance between them.

  A graph cut is the process of partitioning a directed or undirected graph into disjoint sets. The concept of optimality of such cuts is usually introduced by associating an energy to each cut [45].

2.2.7 Feature Based Alignment

Once we have extracted features from images, the next stage in many vision algorithms is to match these features across different images to perform an alignment [26].

An important component of this matching is to verify if the set of matching features is geometrically consistent, e.g. whether the feature displacements can be described by a simple 2D or 3D geometric transformation [34]. More complex models can be also applied to non-rigid or elastic deformations [46].

Given a set of matched feature points and a planar parametric transformation, the solution is usually obtained by using a least squares minimization method [47].

These techniques are usually used in the estimation of camera pose and its internal calibration parameters.

2.2.8 Structure from Motion

This topic is focusing on the process of finding the three-dimensional structure of an object by analysing local motion signals over time.

Structure from motion is related to the kinetic depth effect in perception whereby subjects viewing the shadow cast by a structure in rotation perceive the full three-dimensional structure of the object, whereas when viewing the shadow of a static object they perceive only its two-dimensional projection.

Since the image formation process is not generally invertible, the solution to this problem typically requires to find the correspondence between a set of images and to use this information to reconstruct the 3D object. The general process can be described as follows [48]:

First, correspondence between images is calculated. Therefore features such as corner points (edges with gradients in multiple directions) need to be tracked from one image to the next (see section 2.2.5 and 2.2.7).

Then, knowledge about camera calibration and pose (see section 2.2.1) is used to reconstruct the 3D points by triangulation. The geometrical theory of structure from motion allows projection matrixes and 3D points to be computed simultaneously using only corresponding points in each view.

Other possibility to solve the problem is using assumptions like parallelism and coplanarity constraints instead of image correspondences.

2.2.9 Motion Estimation

Motion estimation, also called image alignment, is related with feature alignment techniques (see section 2.2.7) and is focused on obtaining the motion field corresponding to a sequence of images.

Motion estimation techniques are among the most widely used in Computer Vision. They are used, for example, in image stabilization in cameras, video compression schemes.

- **Block-Matching:** It is conceptually the simplest motion technique based on finding the alignment between two images choosing an error metric and then trying all possible alignments between regions. Pyramids, Fourier Transforms, layering schemes and mathematical models can be used to improve the results. [49].
- **Lucas-Kanade:** The method is a widely used differential method that assumes that the flow is essentially constant in a local neighborhood of the pixel under consideration (the Block-Matching technique uses a similar assumption). Then it works solving the basic optical flow equations for all the pixels in that neighborhood by the least squares criterion. [50].
- **Horn and Schunck:** This algorithm is formulated as an energy functional which is then sought to be minimized. The algorithm assumes smoothness in the flow over the whole image. Thus, it tries to minimize distortions in flow and prefers solutions which show more smoothness. [51].
- **Motion Models:** Estimation can be performed or improved by combining different techniques with the knowledge about the typical dynamics or motion

statistics of the scenes or objects being tracked. An example is the use of parametric motion models such as geometric transformation, and also more complex models such as two dimensional spline controlled by a small number of vertices. [34, 49].

- **Spatio-temporal Filters:** These are physiologically based models based on spatio-temporal energy mechanisms. In these models, oriented or steerable filters are applied to the image [52], in a manner analogous to oriented edge detection.

Tracking

Tracking is the problem of generating an inference about the motion of an object given a sequence of images.

While tracking can be solved using several approaches, including motion estimation techniques and feature matching techniques (see section 2.2.7), some of the most important approaches to tracking use a statistical point of view to formulate the problem as a prediction correction process based on the Bayesian theory. Therefore, the objective is to predict the state of the object at a given frame, having measurements from the previous states of the object.

The main techniques using this approach can be described as follows [37]:

- **Linear Dynamic Models:** These models are mainly based on linear transformations and Gaussian probability densities. For example, if the dynamic model of the object is not known, the new position can be modeled as the previous position plus a Gaussian noise term. This model can be improved by adding knowledge about the movement of the object, such as constant velocity, constant acceleration or periodic movement assumptions. Also higher

order models are possible, representing the actual state in function of several previous states.

- **The Kalman Filter:** Also known as linear quadratic estimation, it uses a linear dynamic model and assumes that the error and probability distribution functions are normal (Gaussian). The algorithm works in a two-step recursive process: In the prediction step, the Kalman filter estimates the current state along with their uncertainties. Once the outcome of the next measurement (corrupted with noise) is observed, these estimates are updated using a weighted average, with more weight being given to estimates with higher certainty. The algorithm can run in real time using only the present input measurements and the previously calculated state.
- **Non-Linear Dynamic Models:** Many natural dynamic models are non-linear with probability models that tend not to be normal. Dealing with these phenomena is difficult and there is not a general solution.

  One of the used techniques in these cases is to linearize the model locally and to assume that everything is normal. Other of the used solutions is the particle filtering technique.
- **Particle Filtering:** Also known as a sequential Monte Carlo method, it is used in tracking to estimate the Bayesian probability distributions which cannot be approximated by Gaussians, such as a multi-modal distribution. The basis of this method is to represent the required posterior density function by a set of random samples with associated weights, a weight representing the quality of that specific particle. The particle filter is recursive and operates in two phases: In the prediction step, the probability distribution of the current state is predicted by updating the previous particles according to a motion model. Then, in the update stage, the weights are re-evaluated based on a likelihood score using the new observations [53].

2.2.10 Stitching

Originated in photogrammetry, this topic is focused in using different images from different parts of a scene to integrate them in a seamless photo-mosaic [47, 54].

Stitching is a problem related to motion estimation, stereo correspondence, shape recovery and feature alignment.

To solve this issue, currently most algorithms use an approach based on aligning the image using a set of features to solve a motion model [55].

Applications using these techniques have been used in a large number of commercial products.

2.2.11 Computational Photography

Computational Photography is focused on using image analysis and processing algorithms in one or more photographs to create images that go beyond the capabilities of traditional imaging systems.

Image stitching can be seen as a special computational photography technique. Other applications using this technique include capturing the full range of brightness in a scene through the use of multiple exposures or algorithms that merge flash and regular images to obtain better exposures [56, 57].

Some of these techniques are now being incorporated directly into digital still cameras.

2.2.12 Stereo Correspondence

Stereo matching is the process of taking two or more images and estimating a 3D model of the scene by finding matching pixels in the images and converting their 2D positions into 3D depths.

The most common application in stereo correspondence is using two frontal cameras to take photos from the same scene and then, aligning the images from the cameras, obtaining a perception of the depth in the scene.

The human vision works with a similar principle, called motion parallax. According to it, the amount of horizontal motion between the two frames (disparity) is inversely proportional to the distance from the observer, and this information can be used to calculate a 3D model.

Image alignment techniques were also discussed in the motion estimation and stitching. In the case of stereo matching, however, we have some additional information available, namely the positions and calibration data for the cameras that took the pictures of the same static scene. This information can be exploited to increase both speed and matching reliability.

The main principles can be explained considering a point X in the real 3D space, being projected simultaneously in two image points x and x' through two camera projection centers C and C'. The points X, C, C', x and x' lie in a plane called epipolar plane. The projection of the rays through x and X, and x' and X are called the epipolar lines of x and x', respectively. This projection is described by epipolar geometry [26].

There are two fundamental approaches to perform stereo matching:

a. A feature based approach, using a set of potentially matchable image locations, such as corners or edges and then solving a model using those correspondences. [58, 59].
b. A dense approach, calculating a correspondence for each pixel, which is a more challenging problem since it is required to infer depth values in textureless regions [60].

2.2.13 3D Shape Recovery

This topic is focused on reconstructing the shape of an object using one or multiple images [26]. Stereo correspondence (see section 2.2.12) can be seen as a kind of shape reconstruction technique. However, a large variety of techniques exist to perform 3D modeling.

Thus, the shading on a surface can provide a lot of information about local surface orientations and overall surface shape. Additionally, texture gradients can provide similar cues on local surface orientation. Focus is another powerful cue to find scene depth, especially when two or more images with different focus settings are used [61].

3D shape recovery can be more efficient and effective if we know something about the objects we are trying to reconstruct.

2.2.14 Texture Recovery

When modeling, after a 3D model of an object of interest has been acquired (see previous section) with the objective of reconstructing the object's surface appearance, the final step is to obtain a texture map describing the visual properties of the surface.

This requires establishing a parameterization for the texture coordinates *(u, v),* as a function of 3D surface position *(x, y, z)*.

Usually, different texture maps are associated to a representation of the surface based on a set of polygonal faces. Then, one or more source images *j* are used to find the perspective projection equations mapping from texture *(u, v)* to the image coordinates (u_j, v_j).

This can be obtained by concatenating the mapping *(u, v)* → *(x,y,z)* with a perspective projection model describing *(x, y, z)* → (u_j, v_j) [62, 63].

An alternative approach is to estimate a complete surface light field for each surface point [64].

2.2.15 Image Based Rendering

In image-based rendering, 3D reconstruction techniques from Computer Vision (see section 2.2.13) are combined with Computer Graphics rendering techniques that use multiple views of a scene to create interactive photo-realistic experiences. [65].

2.2.16 Recognition

Recognition is the higher level task in Computer Vision, and therefore the most challenging one.

The objective of recognition is to interpret the image, extracting knowledge about the objects present in the image, and therefore it is related to the Artificial Intelligence topic of learning.

The task of recognition can be classified as follows:

- **Object Detection:** When we are looking for a specific object. This technique involves scanning an image to determine where a match may occur. If the object we are trying to recognize is rigid, the problem can be solved searching for characteristic feature points and verifying that they align in a geometrically plausible way. In the non-rigid case, deformation models and complex representations of the objects may be required [18].
- **Class Recognition:** This problem involves recognizing instances of extremely varied classes [66]. To solve this problem some techniques rely purely on the presence of features and their relative positions [67], while other techniques are based on finding the constituent parts of the objects and measuring their geometric relationships. In many instances, recognition depends heavily on the context of surrounding objects and scene elements [68].

The general case of recognition is still far from a solution and there is not even any consensus among researchers about when a significant level of performance might be achieved.

A recognition problem can be usually separated in three stages:

1. Segmentation, (see section 2.2.6)
2. Representation, which involves describing the detected objects using meaningful and invariant descriptors which may be compared to other descriptors. With this purpose scale-invariant features (see section 2.2.5) and graphs can be used; also lines, contours, shape or texture descriptors and higher level characteristics such as inverted indexes or frequency vectors.
3. Recognition, that is determining if a detected object belongs to a class or a set of classes of interest. Usually a distance metric should be defined in this task.

Given the extremely rich and complex nature of this topic, several techniques are used in recognition, ranging from machine-learning techniques such as boosting or neural networks; to support vector machines, subspace (*PCA*) models or Bayesian approaches. However, a detailed overview of these techniques is beyond the aim of this thesis.

2.3 Computer Vision Applications

Computer Vision is being used today in a wide variety of real-world applications, which include [13]:

- Optical Character Recognition (*OCR*): ranging from reading handwritten postal codes on letters to Automatic Number Plate Recognition *(ANPR)*.
- Machine inspection: quality assurance using techniques like stereo vision or X-ray vision with specialized illumination to measure tolerances or to look for defects in certain machine parts such as aircraft wings, auto body parts and others.
- Retail: object recognition for automated checkout lanes.
- 3D model building (photogrammetry): fully automated construction of 3D models from aerial photographs. It is used in systems such as Bing Maps.
- Medical imaging: registering pre-operative and intra-operative imagery, performing long-term studies of people's morphology and others.
- Automotive safety: detecting unexpected obstacles such as pedestrians on the street, or identifying road signs. Some of these applications are based on active vision techniques such as *RADAR* or *LIDAR*.
- Match move: merging computer-generated imagery *(CGI)* with live action footage by tracking feature points in the source video to estimate the 3D

camera motion and shape of the environment. Such techniques are widely used in cinema.

- Motion capture using retro-reflective markers viewed from multiple cameras or other vision-based techniques to capture actors for computer animation.
- Surveillance: monitoring for intruders, analysing highway traffic, and monitoring pools for drowning victims.
- Fingerprint recognition and biometrics: for automatic access authentication as well as forensic applications and others.

While the above applications are all extremely important, they mostly pertain to fairly specialized kinds of imagery and narrow domains, and they represent only a small set of current *CV* based technologies.

Currently *CV* applications must be built *ad hoc*, taking in account the parameters of each problem and the specific conditions of each environment.

Life is short and truth works far
and lives long: let us speak the truth
Arthur Schopenhauer

III. Working Hypothesis

Civil Engineering is a discipline that deals with the design, construction, and maintenance of the physical and naturally built environment; including works like roads, bridges, canals, dams and buildings[67].

Civil Engineering is the oldest engineering discipline after military engineering. It can be divided in several sub-areas; including environmental engineering, geotechnical engineering, geophysics, geodesy, control engineering, structural engineering, transportation engineering, earth science, atmospheric sciences, forensic engineering, municipal or urban engineering, water resources engineering, materials engineering, offshore engineering, quantity surveying, coastal engineering, surveying, and construction engineering.

Civil engineers carry out a huge amount of experimentation in laboratories and in design offices to evaluate and compare basic design configurations, to analyse different properties of materials and to select design parameters [30].

A typical Civil Engineering experiment is basically a test or a series of tests in which changes are made to the input variables of a system to observe, measure and identify variations in the output response.

Usually these experiments cannot be conducted directly in the real systems because they are unavailable, because the experiment is destructive or simply because the experimental factors cannot be controlled in the real system.

In these cases, sometimes it is possible to simulate the system in a computer. A computer simulation is based on an abstract model also called computational model. This model is integrated by algorithms and equations used to capture the behavior of the system of interest.

However, many problems are still too complicated to understand in an analytical manner or the current numerical techniques lack real world confirmation. When this is the case, civil engineers conduct experiments using scale models.

Scale models are constructed observing requirements of similitude in order to maintain all the properties of the material or structure to be analysed.

The main advantages of this procedure are that the experiments can be conducted in controlled laboratory conditions with a reduced cost, and they provide information about the process to be studied without using numerical models which may not capture the complexity of the real system.

Typically, the conducted experiments require to measure different physical properties in the studied system to determine its response to the variation of the

input parameters. To this end, specific instrumentation has been used along the years; such as strain gauges or *LVDT* sensors in material testing (Figure 8), or velocity meters and water quality probes in hydrology (Figure 9)

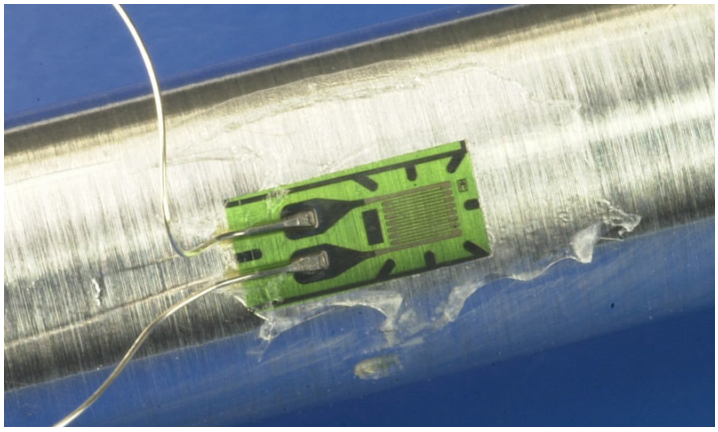

Figure 8: Strain gauge.

Strain gauge attached to a steel bar to measure deformation in a tensile test [55].

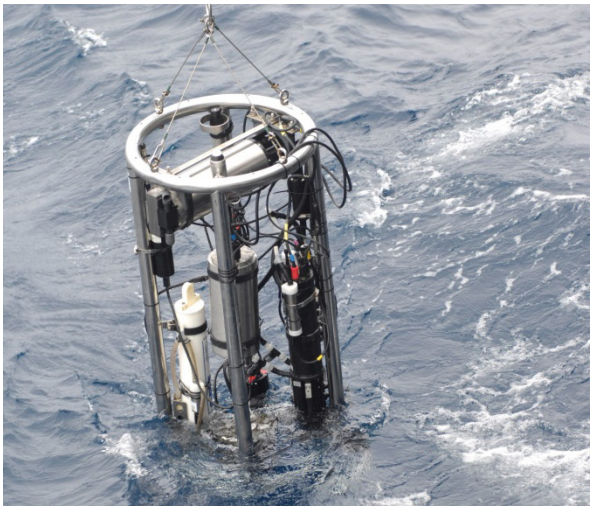

Figure 9: Hydrology sensors.

Use of a velocity meter to measure water speed with other sensors to analyse salinity, temperature, and depth. [56].

Most of these instruments are expensive, provide very specific measurements, and they usually interfere with the studied phenomena. However, their biggest limitation is that they only provide measurements in the position they are fixed.

Computer Vision allows to obtain precise measurements of almost any observable phenomena and may be a more flexible and cheaper alternative to analyse Civil Engineering experiments.

This hypothesis has been already supported by different works like [69, 70], where the use of optical devices was studied to analyse strain in materials, [71], where a Computer Vision system was developed to perform wave flume measurements, or [68], where overflow in channels has been study using image processing.

This thesis assumes the hypothesis of that a Computer Vision system can be developed using a systematic and methodical approach to solve a Civil Engineering problem.

To validate that hypothesis this thesis carries out in the next sections a design methodology, and uses this methodology in two different applications to solve different problems of Civil Engineering.

Caress the detail, the divine detail

Vladimir Nabokov

IV. Methodology

Computer Vision is, as seen in previous chapters, a broad field covering a huge amount of techniques and topics (see section II.2.2).

Therefore, as mentioned in section I.1, *CV* systems and algorithms are usually elaborated *ad hoc*, and the task of defining a methodology using systematic procedures is problematic.

The purpose of a methodology is to establish a framework to create or maintain a system in a deliberate, structured and methodical way. This will not only reduce the cost and time of developing the system, it will also make easier the design process, increase the efficiency of the job and improve the quality of the final system.

A *CV* system in Civil Engineering will have the aim of measuring a real process, usually taking place in controlled laboratory conditions.

Using an intuitive analysis, the first step will be the study of the process of interest and the study of state of the art of similar *CV* applications. Then, the experimental conditions and the acquisition system will be designed and the software will be developed using any standard software methodology.

In this chapter the elements needed in a *CV* system will be defined and the workflow from the problem domain to the final results will be established. Then, the techniques needed in Civil Engineering, including camera calibration and removal of distortion will be discussed. Finally, a general development methodology will be defined; and finally, the design of the experiment, which is probably the most important factor to get accurate results, will be studied.

1 Elements of the System

A *CV* system used in Civil Engineering will basically have the same hardware components than an inspection or industrial system such as the shown in Figure 10. Thus, a general *CV* system in Civil Engineering may have the following elements:

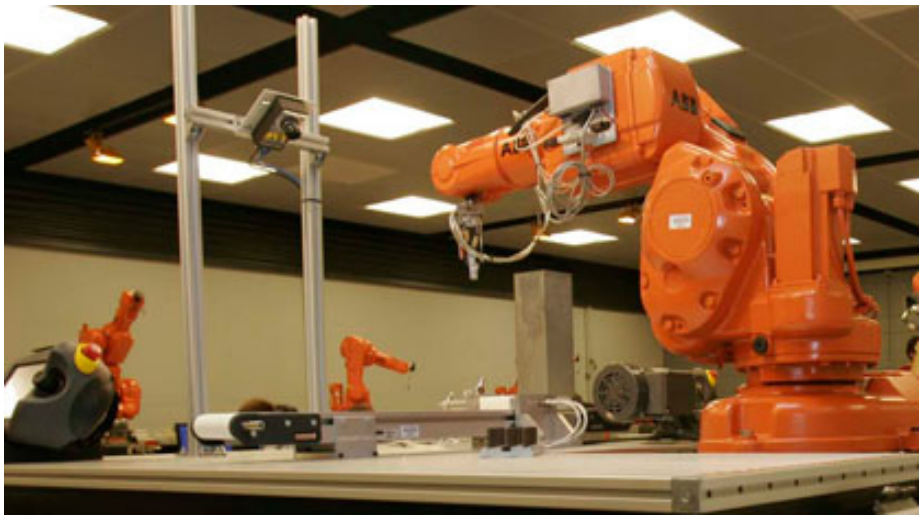

Figure 10: An industrial *CV* system [72].

- **Image Acquisition:** It is a physical device that is sensitive to a band in the electromagnetic energy spectrum and produces an electrical signal output.
- **Storage:** Usually, a *CV* system produce huge amount of data, in the form of images, intermediate results and others, which needs to be stored in different ways. The following storage types will be commonly required:
    - Short-term storage. It will be used during computer processing operations. It is provided usually in the form of computer memory or specialized frame buffer.
    - Mid-term storage. It is used when a relatively fast recall is needed. It will be located on an online support.
    - Archival storage or long-term storage. It is characterized for an infrequent access and it is used to preserve materials in a permanent and durable way. It provides protection against deletion, viruses and disaster.
- **Processing:** It is the element where images will be processed, extracting the required information from the world, and where decisions will be taken. Processing elements can be classified as follows:

- **Hardware Elements:** Currently, several *CV* algorithms can be automatically performed using hardware devices more efficiently than by using software. These processing stages are usually low level and mostly are performed in the image acquisition devices.
- **Software Elements:** In the software elements will be usually located the processing stages of the system, these processes will be implemented in a host computer. However, some of the processing can be also located in the image acquisition system (e.g. smart or programmable cameras).

- **Communication:** It is necessary to access the data from different devices; usually communication is performed with an internet network, using cable modem internet services or *DSL* internet services. Bandwidth may be a bottleneck if image data needs to be accessed from different computers.
- **Display:** An image display device is commonly used to control or supervision processes. Computer monitors are the most common display devices.
- **Actuators:** The Information extracted with the *CV* system may be used in a reactive way to take decisions which will be executed by these devices.

Figure 11, shows an example of a diagram of a typical *CV* system with the elements detailed in this section.

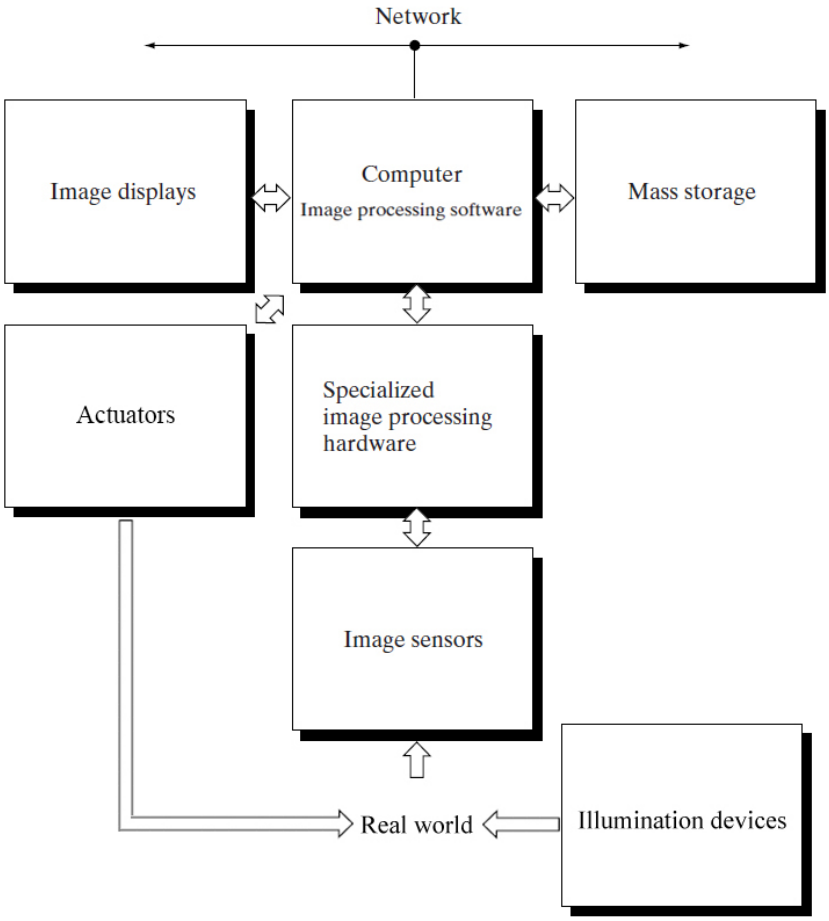

Figure 11: Main components of a *CV* system.

2 Processing Stages of the System

Despite the mentioned variability of techniques and problems in Computer Vision, the systems used to solve a Civil Engineering problem will usually have the same workflow and similar processes.

Figure 12 shows a scheme with the most common processes and levels of processing required. This scheme should be taken only as a reference and most *CV* systems will present only a subset of the defined stages.

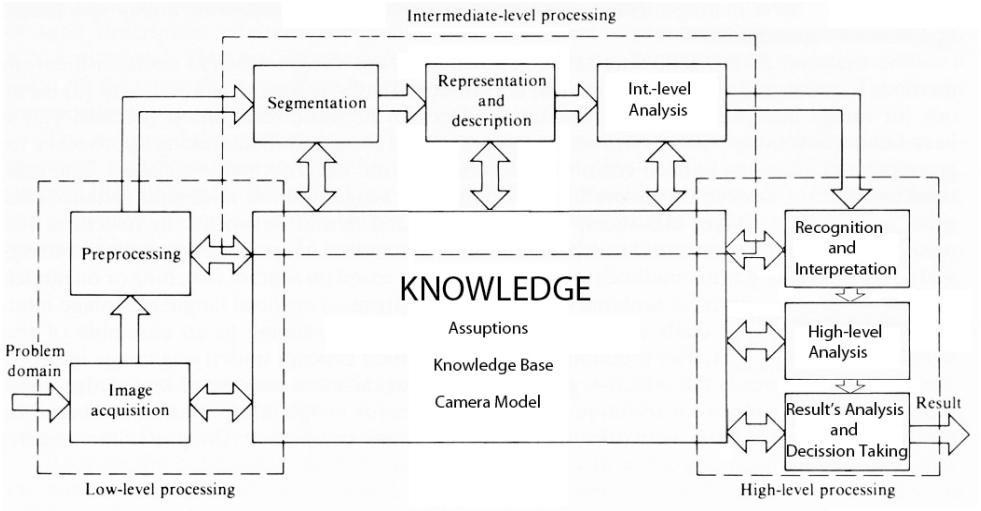

Figure 12: Procedures of a typical *CV* system for Civil Engineering.

According to the level of knowledge used three processing levels can be defined:

- **Low-level Processing:** In this stage, signal or image processing techniques are used and images are not interpreted. Its purpose is to prepare the images for further processing stages.
- **Intermediate-level Processing:** Image processing algorithms are applied in this stage. Simple *a priori* knowledge, usually in the form of assumptions, may be used to solve a concrete task.
- **High-level Processing:** In this level, more ambitious and general tasks are performed. Here, a complex interpretation of the image is used, usually together with *feedback* and Artificial Intelligence methods [4].

Although different problems will involve the use of different algorithms and techniques, there will be a general processing structure shared by most *CV* systems

(see Figure 12). This structure involves a sequence of tasks going from the problem domain to the final results and, although a typical *CV* system will present only a subset of these tasks, this structure defines the workflow and processing stages of almost any *CV* system. These procedures can be defined as follows:

- **Image Acquisition:** The first stage of the vision system is the process of recording the image or sequence of images from the real world and transmitting those images to the computer where they are going to be processed. Image acquisition is a hardware dependent problem.
- **Preprocessing:** After the image has been obtained, various techniques can be applied to the image to reduce the noise or to enhance contrast, sharpness, color, and other parameters. Its finality is to prepare the image for the subsequent processing stages.
- **Segmentation:** After the acquisition and enhancement of the image. The next step is to detect the objects of interest (see section II.2.2.6); *a priori* knowledge of the objects of interest or the background is commonly used in this stage.
- **Representation and Description:** The objects detected in the previous stage are translated into convenient descriptors which can be used by the computer to perform the required operations in the subsequent stages of the system.
- **Intermediate-level Analysis:** The representations from the previous stages are processed; these operations are performed using computational representations of the elements in the image. No high level interpretation processes have been made to this point. Examples of these operations are correlation processes, mathematical morphology operations, and others.
- **Recognition and Interpretation:** In this step, the computational representations of the elements in the image are interpreted and associated (matched) with *a priori* concepts or classes obtained from the real world.

- **High-level Analysis:** This stage involves using real world models provided from the interpretation stage to extract useful information from the object or to perform a *feedback* using meaningful information to improve the output of the system. Examples of these operations can be detecting special characteristics in objects and using high level knowledge to enhance a background modeling or segmentation process.
- **Results Analysis and Decision Taking:** This stage involves processing the results and obtaining conclusions or taking actions based on them.

3 Knowledge

As discussed in section II.2, Computer Vision deals not only with image processing; it is characterized for high level processing and its final purpose is to obtain a meaningful representation from the real world through images.

One of the key concepts in Computer Vision is that an interpretation of the images is necessary in order to obtain a representation of the world. This implies that, to obtain knowledge from images, it is necessary to add knowledge to the system, usually in the form of *a priori* obtained information, but also as a *feedback* from the system.

The knowledge used in the *CV* system is usually a determinant factor to obtain results impossible to achieve only with image processing techniques or, at a minimum, is decisive in the quality of the obtained results.

3.1 Assumptions

Assumptions are domain dependent and usually establish conditions required to find a solution to a problem, either by imposing restrictions to the solution itself or by restricting the search space.

In Civil Engineering and other practical fields, the main requirements are accuracy and robustness of the obtained results.

On the other hand, measurements will be usually taken on controlled conditions, where lightning and camera parameters can be set up by the experimenter.

In this context, it is extremely important to use *a priori* knowledge in the most effective possible way.

Thus, when developing a *CV* system, it will be extremely important to:

- Make strong assumptions about lighting conditions.
- Make strong assumptions about the position of objects.
- Make strong assumptions about the type of objects.

Therefore, using *a priori* knowledge, the task can be significantly simplified and the quality of the obtained results may be improved.

The use of strong assumptions, however, will made decisive the design of the experiment. This stage will be discussed later.

3.2 Knowledge Base

The knowledge base will be formed by previous information about the objects or classes of objects of interest, as well as by the knowledge extracted from the elements in the image by the *CV* system itself.

The knowledge base will be used to associate the elements of the image to classes of objects or concepts from the real world, usually by comparing descriptors of the elements in the image; such as size, color or shape, with information collected in a database.

Therefore, the objects in the image can be represented as instances of classes of objects, obtaining high level information from low or intermediate level information, in other words, interpreting the image and providing the image with a meaning.

The concrete knowledge about the elements of the images, which must be differentiated from the *a priori* and abstract knowledge from classes and concepts from the real world, can be used also to analyse the image.

Therefore, in a process called *feedback* the results obtained only with a low level analysis of the visual information can be improved. But also further high level stages of processing can be performed.

As an example, after a segmentation process, two elements of the image are separated from the background. Using a descriptive representation of the objects, and comparing these descriptions with the information in the database, they are classified as two categories.

This information can be used to improve the segmentation process: one object can be a constitutive part of the other so they can be merged, or it can be a non-interesting object and therefore it can be added to the background.

That information, however, can be used also to perform a forward analysis in the two objects; we can be interested, for example, in using different models to describe the shape of the two objects and to find some parameters of interest.

3.3 Camera Model

The camera model describes how a point is projected from the real 3D space to the coordinates in the image as expressed as follows.

$$(x_i, y_i) = f(X, Y, Z) \quad (1)$$

Where *(X, Y, Z)* are the coordinates in the real world, (x_i, y_i) the coordinates in the image and *f* the transformation.

The calculus of the camera model can be divided in three stages:

1. Camera pose estimation, which is estimating the rotation and position of the camera. It describes a transformation from the object coordinate system to the camera coordinate system.
2. Camera projection model, which defines a transformation from the camera coordinate system to the image coordinate system in an ideal camera.
3. Distortion removal, which describes the deviation from the actual position in the image and the position predicted with the camera projection model.

3.3.1 Camera Pose Estimation

The camera model assumes that the camera is first located at the origin of the world coordinates and then translated and rotated [8].

The problem of estimating the camera pose is shown in Figure 13, and the numeric formulation is expressed in (2).

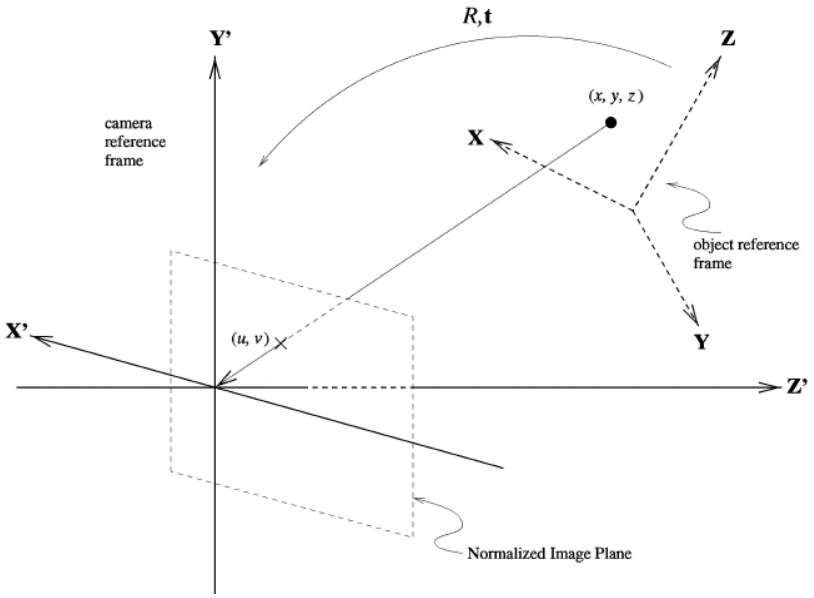

Figure 13: Camera pose [73].

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = R_{3x3} \times \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + T_{1x3}$$

$$R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \quad T = \begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix} \tag{2}$$

Where *(X, Y, Z)* are the coordinates of the point in the real world, and (X_c, Y_c, Z_c) in the camera coordinate system (or reference coordinate system).

The transformation is modeled by a set of parameters called the extrinsic parameters of the camera.

These parameters are formed by *R*, the rotation matrix, and *T*, the translation matrix. They are scene dependent and they are used to describe the camera motion around a static scene; or vice versa, a rigid motion of an object in front of still camera.

3.3.2 Camera Projection Model

The camera projection model describes how a point in the camera space is projected on the image plane trough an optical system. The most used projection model is the pin-hole camera model shown in Figure 14 [27].

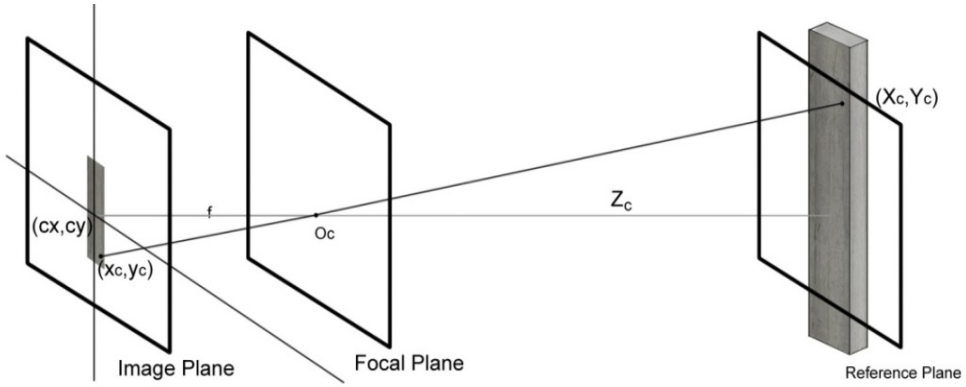

Figure 14: Pin-hole projective model.

A point is projected from a reference plane onto the camera focal plane and to the sensor (image plane) where the image is formed.

The pin-hole model defines the transformation of a point from camera space coordinates *(Xc, Yc, Zc)* to image coordinates (x_i, y_i), as follows:

$$\begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = M_{3x3} \times \begin{bmatrix} X_c/Z_c \\ Y_c/Z_c \\ 1 \end{bmatrix} \qquad M = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 0 \end{bmatrix} \quad (3)$$

Where M is the transformation matrix, $f(f_x, f_y)$ represents the focal length, the distance from the lens to the camera sensor, and $c(c_x, c_y)$ determines the optical center, establishing as reference the image coordinates where a point is projected through the center of the lens Oc.

The matrix M is called the matrix of the extrinsic parameters of the camera, it does not depend on the scene viewed and, once estimated, can be re-used as long as the focal length is fixed.

3.3.3 Distortion Removal

In practice, due to small imperfections in the lens and other factors, some distortions are integrated into the image. These distortions can be modeled using the following parametric equations [74]:

$$\begin{aligned} dr_x(x) &= xk_1r^2 + xk_2r^4 + xk_3r^6 \\ dr_y(y) &= yk_1r^2 + yk_2r^4 + yk_3r^6 \\ dt_x(x,y) &= k_3\left(r^2 + 2x^2\right) + 2k_4xy \\ dt_y(x,y) &= 2k_3xy + k_4\left(r^2 + 2y^2\right) \end{aligned} \quad (4)$$

Where x and y are spatial coordinates, r is the distance to the lens optical center, dr is the radial distortion, dt is the tangential distortion and k_i are the distortion parameters to be determined.

The distortion coefficients do not depend on the scene viewed, thus they also belong to the intrinsic camera parameters. And they remain the same regardless of the captured image resolution.

3.3.4 Solving the Model

The complete camera model can be described as shown in the following equations:

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = R_{3x3} \times \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + T_{1x3}$$

$$\begin{bmatrix} x_c \\ y_c \end{bmatrix} = \begin{bmatrix} X_c / Z_c \\ Y_c / Z_c \end{bmatrix} \qquad (5)$$

$$\begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = M_{3x3} \times \begin{bmatrix} x_c + dr_x(x_c) + dt_x(x_c, y_c) \\ y_c + dr_y(y_c) + dt_y(x_c, y_c) \\ 1 \end{bmatrix}$$

Where it is necessary to estimate 21 parameters and 12 of these parameters are scene dependent.

To solve this model, the calibration technique proposed originally by Zhang [27] is used. This technique consist in comparing the real and the observed geometry of a calibration pattern of known dimensions and then using a optimization technique to minimize the reprojection error, i.e. the total sum of squared distances between the observed feature points and the projected object points (using the current estimates for camera parameters and the poses).

Figure 15 shows an example of calibration, using a checkerboard calibration pattern.

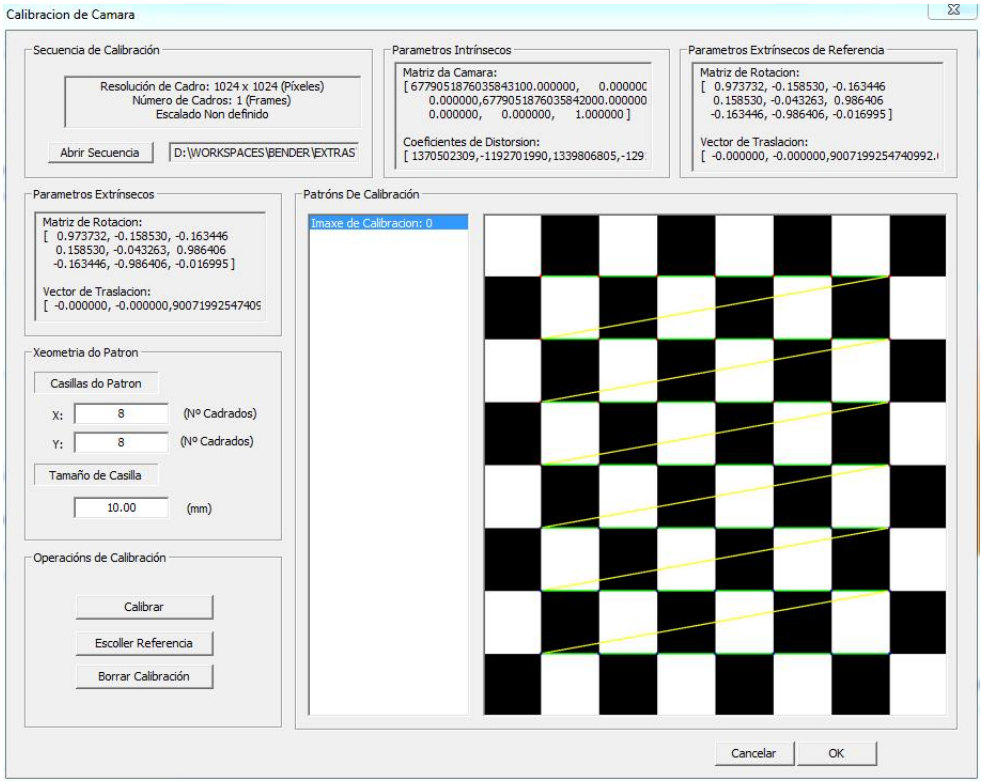

Figure 15: Camera calibration.

Capture of an application to calibrate a camera developed within this thesis.

3.3.5 Multi-Camera Scenario

The model exposed so far is defined using a single camera, however, several *CV* applications would require a multi-camera system. Being the most common case, a stereo system such as the used for stereo correspondence (see section II.2.2.12).

In this context for a given number of cameras, the calibration problem consists in obtaining the intrinsic (calibration matrix) and extrinsic (relative rotation and translation) parameters of every camera from a set of images.

A multi-camera system, can be calibrated just by calibrating each camera independently, however, it is possible to take advantage of using a technique to calibrate the whole system at the same time, and using a calibration pattern in the same real space to every camera.

For example, if we have a stereo camera, the relative position and orientation of the 2 cameras is fixed, and if we compute poses of an object relative to the fist camera and to the second camera, obviously, those poses will relate to each other, if the position and orientation of the second camera relative to the first camera is known. This is expressed in (6):

$$R_2 = R \times R_1 T_2 = R \times T_1 + T \quad (6)$$

Being R_1, R_2, T_1 and T_2 the rotation and translation parameters of the cameras 1 and 2 respectively, R the rotation matrix between the 1st and 2nd cameras' coordinate systems and T the correspondent translation matrix.

Furthermore, using this approach, it is possible to compute the rotation matrixes for each camera that make both camera image planes the same plane. Consequently, that makes all the epipolar lines parallel and thus simplifies the dense stereo correspondence problem [75].

This can be generalized to a system with any number of cameras.

4 Development Methodology

A development methodology refers to the framework that is used to structure, plan, and control the process of developing the *CV* system.

In this context a *CV* system can be considered as a special kind of information system, where it is not only necessary to design and implement a software system, but also to design the experimental conditions and the hardware structure for image acquisition, storage and for the communication parts of the system.

It must be taken into account that one system development methodology is not necessarily suitable for use by all projects. Different methodologies are best suited for specific kinds of projects, based on various technical, organizational, project and team considerations.

In this context, we may use the general software development methodology and we may include the special characteristics of *CV* systems used in Civil Engineering, which can be enumerated as follows:

- The initial requirements of the system are usually well established, although additional functionalities may be added incrementally. Modifications to initial requirements can arise along the project.
- It is usually necessary to explore different techniques and software solutions before reaching the final system. Specific problems will be usually detected in this stage. Therefore, it will be necessary a strong support for iterative development and fast prototyping.
- Innovation and flexible designs are usually necessary to solve new problems.
- The developed software and the required knowledge is strong dependent from the acquisition system and the experimental conditions such as illumination. These elements of the system are sometimes expensive, and difficult to modify.
- An expert in the Civil Engineering field should be involved throughout the development process. The participation of the expert is especially necessary to design the experiment and to define the initial requirements of the system.

- Team members and project manager should be experienced, and team composition should be stable.
- Evaluating alternatives and resolving risks is important through the entire cycle.
- Development methodology should be flexible to be adapted to each project.

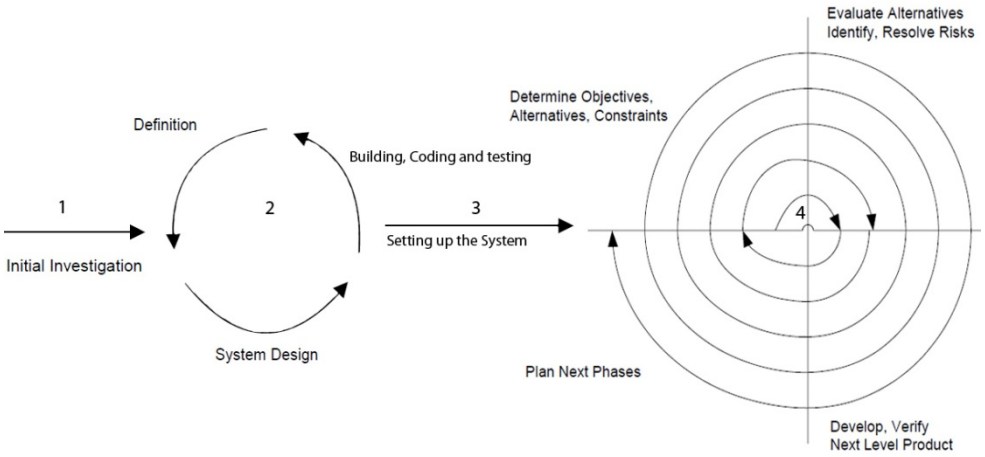

Figure 16: Development methodology.

In this thesis, to handle the characteristics of *CV* systems in Civil Engineering, a spiral methodology is proposed, modified to include linear and prototyping stages.

Figure 16 shows a diagram with the main with the proposed methodology and its main steps are explained as follows.

4.1 Initial Investigation

This will be the first stage of development, where the project is conceived. The aim of this stage is to establish the viability and size of the project. In this stage, the characteristics of the measured process will be studied together with the state of the art techniques to analyse similar processes in Computer Vision.

4.2 Prototyping

This is the most critical stage of the development and it is a research stage. Its purpose is to define the initial requirements of the system, to design the experimental conditions such as illumination and to design a prototype with the acquisition, storage elements and the other components required to obtain the images which will be analysed.

Since the processing techniques and the collected images are codependent; while the hardware part of the system is being designed and build, different techniques and algorithms will be tested to detect the required changes for the next stage, as well to set the starting point to develop the required software.

This stage follows a methodological approach called prototyping. Prototyping is a fast iterative methodology based on constructing small versions of the system. It is ideal to test different approaches, identify problems, establish requirements, define the experimental conditions and it allows the software engineer some insight into the accuracy of initial project estimates.

The prototyping stage can be divided in the following stages:

- **Definition**: In this stage, the inputs, the requirements and the desired outputs will be studied and identified.
- **System Design:** Attending to the defined requisites**,** a low scale prototype of the system will be designed. This will include a plan of the experimental conditions and the hardware components of the system. In this stage, a conventional low cost and low scale acquisition and illumination system will be used to obtain preliminary images.

- **Building, Coding and Testing:** The designed prototype will be build, and it will be used to obtain the preliminary images. Then, images will be analysed using standard techniques or libraries and, developing a software with minimum functionalities. The purpose of this phase is not only to test different techniques and approaches, but also to detect problems and weakness in the obtained images which may lead to modify the design of the system. For example obtaining requirements about the camera type, the camera position, the illumination techniques, lenses, filters, storage or communication requirements… etc.

  In this stage usually only low and mid processing stages will be tested and details such as interfaces and efficiency can typically be ignored.

After each cycle of the prototyping stage, a *feedback* will be established to improve the prototype, modify the specifications and negotiate the scope of the product. If changes are needed, another cycle may be necessary.

The purpose is to reduce the cost and development time of the final system by early identifying the unpredicted characteristics and problems.

With this approach, the main risks of the project will be detected and handled in a low-scale version of the system, preventing costly modifications in the final system.

It is necessary, however, to study the problems which may arise due to the scalability of the system, since this problems may not be apparent in a low scale prototype and may not be detected in this stage.

4.3 Setting up the System

With the knowledge obtained in the previous stage, the requirements of the system will be defined and the experiments will be carefully designed, while the software application will be usually discarded.

In this point a functional prototype of the hardware part of the system will be build. This will normally include specific image acquisition components and costly equipment which will be difficult to modify, while communications, storage and visualization components may be not required in its final form.

Some other aspects, such as illumination conditions and minor changes in camera position or background, will be still refined in later stages of the system.

4.4 Spiral Development

In this point, the hardware design of the system, the initial requirements and the experimental conditions will stable and subject only to minor and non-costly modifications.

In this stage of the system, the *CV* software will be developed with a standard spiral development cycle.

Spiral methodology tries to combine advantages of top-down and bottom-up concepts by minimizing the project risks in a flexible methodology. It is based on the combination of the idea of prototyping with the systematic, controlled aspects of linear models. It allows for incremental releases of the product and incremental refinement through each time around the spiral. The spiral model also explicitly includes risk management.

The spiral stage can be divided in the following substages:

- Determine the objectives, alternatives, and constraints on the new iteration
- Evaluate alternatives and identify and resolve risk issues.
- Develop and verify the product for this iteration.
- Plan the next iteration.

The spiral cycle has the advantage of being a meta-model which can be adapted to different projects by using other models.

5 Design of the Experiments

The design of the experiment is probably the most important and underrated aspect in *CV* systems.

The design of the experiment basically establish the recording conditions; defining the type of illumination, the position and type of the cameras and also different aspects such as the filters or type of lens used to record the assay.

These aspects determine directly the quality of the processed images, and can be used to emphasize the aspects of interest in the scene or to mitigate the undesired elements.

Therefore, the design of the experiment may have a direct influence in the quality and accuracy of the results, and also in the complexity of the required analysis techniques. Furthermore, the analysis of a scene can be trivial with an adequate design, and impossible with an inadequate one.

5.1 The Acquisition System

The acquisition system is formed by the hardware elements where the light is captured and the image formed and digitalized, some initial image processing techniques may be performed in these elements.

Typically, the image acquisition system is formed by the following components.

5.1.1 Camera

The function of the camera is to capture the light forming the image to transmit it to an electronic system where it can be visualized, processed, interpreted or stored.

Currently, most cameras are digital. They capture the light projected through an optical system formed by a set of lens. The image is formed in a sensor with millions of photo detectors.

Cameras used in Computer Vision usually require specific characteristics, such as shot control and complete control of the used times and signals, such as sensibility or shutter speed.

According to the geometry of the sensors, cameras can be classified as follows:

- **Matrix Cameras:** A matrix camera has a sensor formed by a 2D matrix of photodetectors. This is the most common type of camera.
- **Linear Cameras:** These cameras have a linear sensor where the image is formed line by line by moving the camera respect to the object of interest, or moving the object of interest respect to the camera. These types of cameras are used mostly in processes of web inspection, applied to continuous materials such as paper or textiles.

Nowadays, matrix cameras are by far the most used in *CV* systems, while linear cameras are used only in very specific applications.

Cameras can also be classified according to the technology of their sensors as follows:

- **CCD Cameras:** CCD (Charged Coupled Device) sensors contain an array/matrix of photodetectors (pixel sensors) called p-doped MOS capacitors which allow the conversion of incoming photons into electron charges at the semiconductor-oxide interface. They are characterized for the use of displacement registers to transfer the image. CCD image sensors are widely used in professional, medical, and scientific applications.
- **CMOS Cameras:** CMOS (Complementary Oxide Semiconductor) sensors, are formed by an array/matrix of photodetectors called CMOS APS (Active Pixel Sensor), consisting on pinned photodiode in a small circuit. The CMOS sensors can use two different techniques to transfer the image: the global shutter technique (every pixel of the sensor is sensible to the light at the same time) and the rolling shutter technique (where a displacing window is used); being this last technique cheaper with the drawback of creating distortions in the presence of movement. Nowadays, the CMOS sensors have been popularized in a wide range of applications and stand for the best solution for high-speed applications.

The CMOS sensor has the advantage over CCD sensor of being less expensive, faster and immune to the blooming effect (where a light source has overloaded the sensitivity of the sensor, causing the sensor to bleed the light source onto other pixels).

However, the CCD sensor has the advantages of a greater fill factor (the percentage of the area of the pixel sensible to the light), since in a CMOS sensor

only a small part of the area is occupied by the photodiode, and a greater dynamic range (the ratio between the maximum and minimum measurable light intensities): Additionally, the CMOS sensors present some issues like the fixed pattern noise.

To capture color, most CCD and CMOS sensors use a pattern of filters called Bayern mask, with these technique each square of four pixels has one filtered red, one blue and two green; where the color values in the holes are interpolated to obtain a full resolution color image. Nowadays some cameras use a beam splitter prism and three different sensors to capture color; however, this approach is usually too much expensive.

To measure the quality of a digital camera, the following parameters can be considered [76]:

- **Resolution:** The resolution describes the detail which can be hold by an image. It is determined by the number of photodetectors and pixel locations on the sensor, and it is usually measured in megapixels. The effective resolution however, also depends on the ability of the lens to match the sensor resolution.
- **Color accuracy:** Conventional sensors using a color filter array have only one photodiode per pixel location and will display some color inaccuracies around the edges because the missing pixels in each color channel are estimated. Other factors such as lens chromatic aberrations have big influence in this aspect.
- **Dynamic Range:** The ratio between the maximum and minimum measurable light intensities is an important factor to measure quality. One of the most important aspects regarding the dynamic range is the size of the photodetectors, being the dynamic range directly proportional to this size.

- **Noise:** Is an unwanted random addition to the image signal. The quality of the sensor and the size of its pixel locations have a great impact on noise and how it changes with increasing sensitivity.

It can be stated, therefore, that the quality of a camera relies on a compromise between the number of photodetectors and its size, being these two opposite philosophies for a defined sensor size. Generally, it is assumed by experts that it is more important to have less resolution but better image quality, although common belief considers image quality as being directly proportional to the megapixels of the sensor.

Video Cameras

Basically a video camera and a photographic camera share the same principles. A video camera can be simulated by a photographic camera operating in a continuous shooting mode, and nowadays, most photographic cameras include a video mode.

The most important differences are therefore, the video compression formats and algorithms and the video capturing and transmission techniques and signals.

Video signals can be interpreted in two different ways:

- **Progressive Scan:** With progressive scan, an image is captured, transmitted, and displayed in a path similar to text on a page: line by line, from top to bottom.
- **Interlaced Scan:** The interlaced signal contains two fields of a video frame captured at two different times, having each field only half of the lines needed to make a complete picture. This results in an effective doubling of time resolution as compared with non-interlaced footage. However, to analyse an

interlaced video in a *CV* system, the image must be deinterlaced and the artifacts in the interlaced signal cannot be completely eliminated because some information is lost between frames.

Special Cameras

- **High Speed Cameras:** Standard video camera records at a speed of 25 or 30 frames per second and common commercial photographic cameras can operate in a continuous mode at no more than 10 frames per second.

  However, many fast processes require a higher speed to be recorded properly. There are some sensor features that can be used in high speed applications such as the possibility of some interlaced sensors of using the half of the vertical resolution doubling the speed or the binning mode (the sensor reads two or more pixels at the same time, adding the resulting charge for each pixel) which reduces the resolution of the sensor gaining speed and signal to noise ratio.

  However, for applications where the speed is critical, there are specialized high speed cameras which can record faster than 10.000 frames per second using a megapixel resolution.

  In these applications a critical aspect is the shutter speed, which defines the exposition time, the time used for the sensor to capture the light to form the image. Increasing the shutter speed means reducing the light captured by the camera, so the luminosity of the lens and the intensity of the illumination became critical aspects.

  The speed of the camera is also limited for the speed of the data bus, the memory or buffers and the writing speed of the image, since a high speed camera can generate more than 10 gigabytes of image data per second.

- **Infrared Cameras:** An infrared camera is a device that forms an image using infrared radiation. Infrared cameras operate in wavelengths of the electromagnetic spectrum as long as 14,000 nm. Infrared cameras are also called thermographic cameras, since they can measure the changes in the temperature of the bodies, given that all objects emit a certain amount of black body radiation as a function of their temperatures. The higher an object's temperature is, the more infrared radiation is emitted as black-body radiation.
- **Multispectral Cameras:** Multispectral cameras capture image data at specific frequencies across the electromagnetic spectrum. The wavelengths may be separated by filters or by the use of instruments that are sensitive to particular wavelengths, including light from frequencies beyond the visible light range, such as infrared. Spectral imaging can allow extraction of additional information the human eye fails to capture with its receptors for red, green and blue. It was originally developed for space-based imaging.
- **Submarine Cameras:** *CV* applications in underwater environments are specially complicated, as the camera must be isolated from water (using special impermeable cases) and it must face several issues which affect the quality of the obtained image such as visibility conditions and effects like absorption of light, reflection of light, bending of light, denser medium, and scattering of light [77].
- **Intelligent Cameras:** They are cameras which integrate special image processing software to perform different *CV* functions. Currently, almost every commercial camera can be included in this category, since they all use an image processor and are able to apply different processing techniques such as face detection or noise removal.
- **3D Laser Cameras:** These cameras allow performing 3D measurements using laser structured illumination (see section 5.2) with a software processing unit to analyse the light patterns projected in the objects.

- **3D Time of Flight Cameras:** Time-of-Flight (*TOF*) cameras use an active illumination unit to send out intensity-modulated light in the near-infrared range. If the light hits an object it is reflected back to the camera making possible to compute the distance of the sensor to the illuminated object for each sensor pixel.

Frame Grabbers and Sync to Cam

One important aspect in a *CV* application is to obtain a picture or a sequence of pictures in the desired moment or with the desired frequency. This objective can be achieved in different ways depending on the requisites of the application.

If the events of interest can be controlled by user or the scene of interest is static, a commercial camera can be used and the images taken manually.

In some applications it would be enough to use a commercial camera to record the desired process or experiment in video mode or using a continuous shot operation.

However, in the most exigent applications special cameras may be required using a special device called frame grabber. A frame grabber is an electronic device that captures individual, digital still frames from an analog video signal or a digital video stream.

The frame grabbers can be connected to a camera or set of cameras to control and synchronize the acquisition process and convert the output from the acquisition devices into data to be used in a computer.

Frame grabbers can be used therefore, to control and synchronize cameras in a precise time using different trigger and clock signals, which can be controlled by

computer and allow video frames to be displayed, stored or transmitted in raw or compressed digital form.

They can also do different functions; such as digitalize analogical images, use different preprocessing techniques, or provide controlling signals to other elements of the system.

Nowadays, some cameras can be connected directly to the computer by different ports such as FireWire, USB or Gigabyte Ethernet and do not require frame grabbers.

Additionally, if a high precision control of the acquisition time is required, it is necessary to use a camera implementing the capture in asynchronous reset technique, to avoid the waiting time for the synchronous signal.

5.1.2 Lenses

Lenses are used to transmit the light to the camera sensor. A lens is in practice made up of a set of mobile and fixed optical lens elements which interact to transmit the light, allowing adjusting the zoom and focus and minimising optical aberrations.

Figure 17: Camera lens [78].

The two fundamental parameters of an optical lens are the focal length and the maximum aperture. The lens focal length determines the magnification of the image projected onto the image plane and the aperture the light intensity of that image.

For a given photographic system the focal length determines the angle of view; short focal lengths giving a wider field of view than longer focal length lenses.

A wider aperture allows using a faster shutter speed for the same exposure, which means that the amount of light captured by a lens is proportional to the area of aperture. The area of aperture is given by the following expression:

$$\text{Area} = \pi\left(\frac{f}{2N}\right) \quad (7)$$

Being f the focal length and N the focal ratio.

The focal length can be seen as the distance from the lens to the camera sensor. It describes the ability of the optical system to focus light, and it is formally defined

as the distance over which initially collimated rays are brought into focus when focusing infinity.

The focal ratio expresses the maximum usable aperture of the lens; it is defined as the lens focal length divided by the effective aperture, resulting a dimensionless number also called f-number.

As a general rule, in an optical system the larger the aperture (lower focal ratios), the higher the light intensity at the focal plane. On the other hand, smaller apertures (higher focal ratios) increase the Depth of Field (*DOF*) allowing objects at a wide range of distances to be all in focus at the same time.

A zoom lens can vary its focal length to amplify the elements of the image (higher focal length) or to reduce the magnification gaining perspective (lower focal length).

Augmenting the zoom means augmenting the focal length, reducing the maximum usable aperture, reducing the maximum luminosity and shortening the depth of field.

When selecting a lens for a *CV* system, the necessary focal length should be calculated using (8) [79]*:*

$$f = \frac{S_S \times d}{O_S} \quad (8)$$

Being *f* the focal length, S_S the sensor size, O_S the object size, and *d* the distance to the object.

To calculate the camera's angle of view α, the following rule can be used [71]:

$$\alpha = 2 \times \arctan\left(\frac{S_S}{f}\right) \approx \frac{180 \times S_S}{\pi \times f} \quad (9)$$

Being f the focal length and S_S the sensor size.

The quality of the lenses can be measured with the following parameters:

- **Modulation Transfer Function (*MTF*):** It quantifies how well a subject's regional brightness variations are preserved when they pass through a camera lens. It is usually expressed in function of line pairs per unit length. The concept of line pairs expresses alternating black and white lines, and the main idea is that beyond the resolution of the lens these lines will be no longer distinguishable. An *MTF* of 1.0 represents perfect contrast preservation, whereas values less than this mean that contrast is being lost. This resolution limit is an unavoidable barrier in any lens; it only depends on the camera lens aperture and is unrelated to the number of megapixels.
- **Relative Illumination Uniformity:** All images from lenses vary in intensity from the center to the edge of the image. The center of the image is brighter than the edges of the image. There are two primary factors involved in this non-uniformity of illumination. First, there is a natural decrease in illumination from the center to the edge of the image circle. This varies with the fourth power of the cosine of the field angle. Second, there is often some mechanical *vignetting* of the image within the lens. It varies in function of the used aperture.
- **Distortion:** The term distortion refers to a change in the geometric representation of an object in the image due plane to imperfections in the shape of the lenses.

- **Chromatic Aberration:** Chromatic aberrations are defects in an imaging system caused by the fact that different wavelengths or colors of light are refracted by different amounts.

5.2 Illumination

Light may be the most relevant aspect in a *CV* system. Without an appropriate illumination, the task of vision is impossible. Furthermore, it has been stated that the 90% of the success of any machine vision application is through proper lighting

The light conditions determine the typo, position, angle and intensity of the beams incident in the object of interest and which will be then registered by the acquisition device.

5.2.1 Light properties

Light is a part of the electromagnetic radiation which exhibits properties of both waves and particles. It travels with a speed approximate of *299.796.000 m/s* and is expressed in function of its wavelength and frequency.

Figure 18 shows the structure of the electromagnetic spectrum.

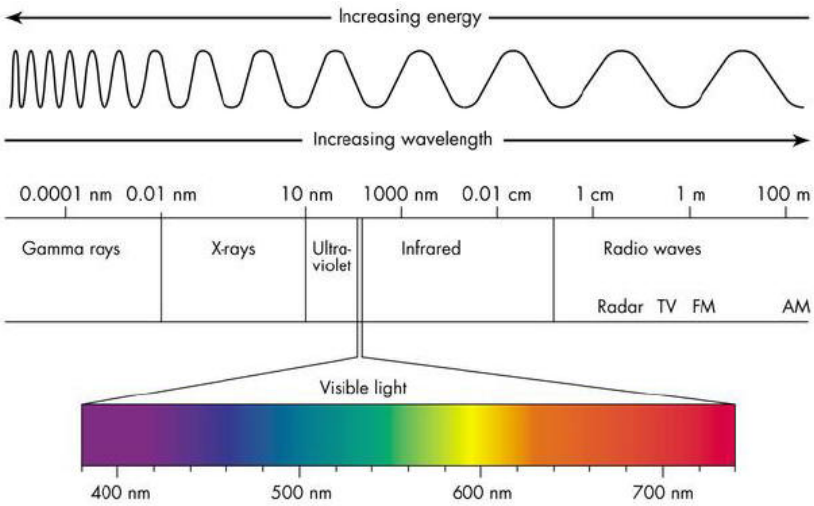

Figure 18: Electromagnetic spectrum [80].

When the light finds an obstacle, it collides with its surface and a part of the light is reflected, other part is absorbed by the object and part passes through the object in a phenomenon called transmission, this process is detailed in Figure 19.

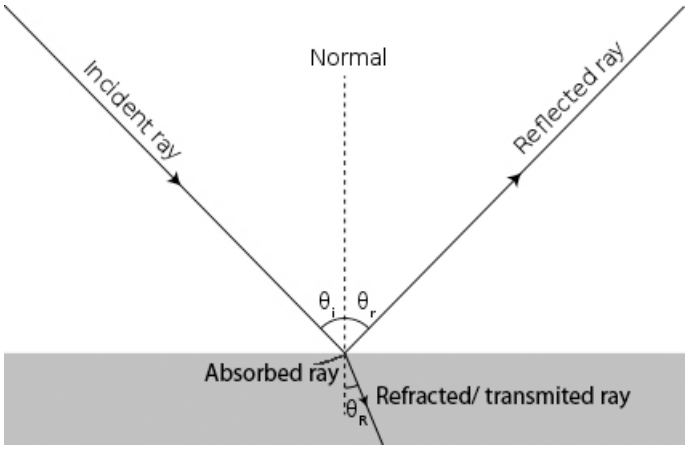

Figure 19: Interaction of light and matter.

Images are formed by the reflected light received and coded by a camera. Therefore, the quality of the image and the accuracy of the image processing

system depend directly on the light received by the object and it is possible to use the properties of light to solve a *CV* problem.

To understand how the properties of the illumination source affect the perceived object, it must be taken into account that there are two types of reflection. The specular reflection, showed in Figure 19, which is reflected in the angle θ_r equal to the angle of the incident ray θ_i. However, there is also a diffused reflection due to a part of the light scattered in all directions. The different types of reflection are shown in Figure 20.

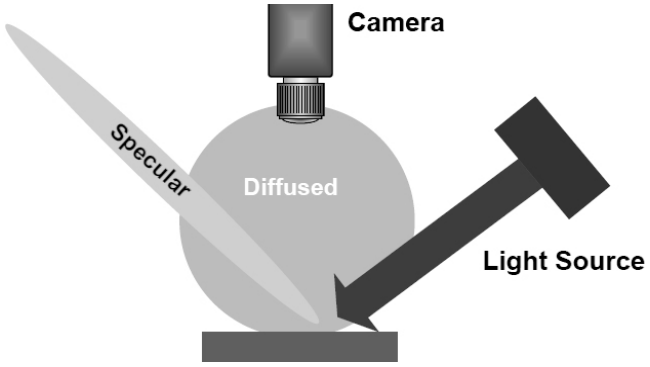

Figure 20: Reflection of light.

5.2.2 Ambient light

The presence of ambient light input can have a negative impact on the quality and consistency of the *CV* system, particularly when using a multi-spectral source, such as white light. The most common ambient contributors are overhead lights and sunlight, and also occasional interferences from other *CV* applications.

There are 3 active methods for dealing with ambient light [81]:

- **High-power Strobing with Short Duration Pulses:** This technique simply overwhelms and washes out the ambient contribution, but has disadvantages in

ergonomics, cost, implementation effort, and not all sources can be strobed, e.g. fluorescent.

- **Pass Filters:** This technique blocks the light received by the camera except for a narrow wavelength which is adjusted to the source of light of interest.
- **Physical Enclosures:** This technique is used when strobing cannot be employed and multi-spectral white light is necessary for accurate color reproduction and balance. In this circumstance pass filter is ineffective, as it will block a major portion of the white light contribution.

There are exceptions to this rule, however. For example, a 700 nm short pass filter, otherwise known as an *IR* blocker, is standard in color cameras because *IR* content can alter the color accuracy and balance, particularly of the green channel.

Figure 21 illustrates how the use of a pass filter can block ambient light very effectively, particularly when the light of interest is low yield fluorescence.

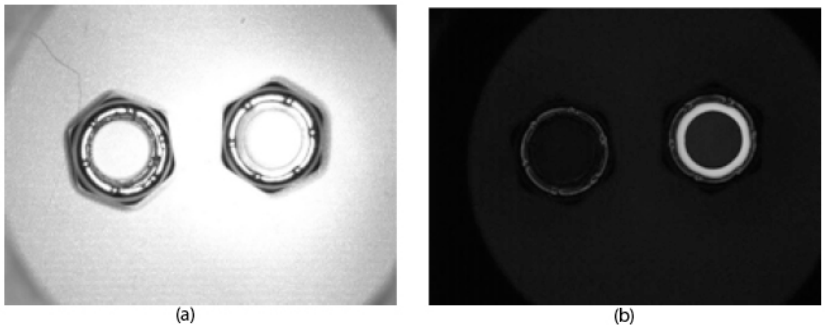

Figure 21: Blocking ambient light.

(a) Image with UV ring light, but flooded with red 660nm ambient light. (b) Same lightning, with a 510 nm short pass filter installed on the camera lens. Blocking the red ambient light and allowing the blue 450 nm light to pass [81].

5.2.3 Illumination techniques

Nowadays exist several light sources and techniques of illumination, with different properties and appropriated for different applications.

The most common lightning sources used in Computer Vision are:

- **Fluorescent:** It is a gas-discharge lamp that uses electricity to excite low pressure mercury vapor. The excited mercury atoms produce short-wave ultraviolet light that then causes a phosphor to fluoresce, producing visible light. It is the most cost-effective for large area lighting deployment.
- **Mercury Lights:** A mercury-vapor lamp is a gas discharge lamp that uses an electric arc through high pressure vaporized mercury and argon to produce light. Mercury vapor lamps are more energy efficient than incandescent and most fluorescent lights.
- **High Pressure Sodium Lights:** As a gas discharge lamp, similar to mercury lamps, that uses sodium (with mercury) in an excited state to produce light using a electric arc. Sodium vapor lamps cause less light pollution than mercury-vapor lamps. They are used in areas where good color rendering is important, or desired.
- **Metal Halide Lights:** At is a variation of mercury lamps, in these lamps light is produced by an electric arc through a gaseous mixture of high pressure vaporized mercury, argon and a variety of metal halides. Metal-halide lamps have about twice the efficiency of mercury vapor lights and 3 to 5 times that of incandescent lights. Metal halide is often used in microscopy because it has many discrete wavelength peaks, which complements the use of filters for fluorescence studies.
- **Xenon Lights:** They are actually metal halide lamps that contain xenon gas. The xenon gas allows the lamps to produce minimally adequate light

immediately upon power up, and accelerates the lamps' run-up time. A xenon source is useful for applications requiring a very bright, strobed light.

- **Incandescent Lights:** They produce light by heating a filament wire to a high temperature until it glows. Currently they offer poor energy efficiency and life.
- **Quartz Halogen Lights:** Is an incandescent lamp that has a small amount of a halogen such as iodine or bromine added. The combination of the halogen gas and the tungsten filament produces a halogen cycle chemical reaction which redeposits evaporated tungsten back on the filament, increasing its life and maintaining the clarity of the envelope.
- ***LED* Lights:** An illumination source based on Light Emitting Diodes (*LED*). Nowadays *LED* technology outperforms in stability, intensity, and cost-effectiveness.
- **Fiber Optics:** Is a flexible, transparent fiber made of a pure glass (silica) to transmit light between the two ends of the fiber. They are used with *LED* or halogen light sources.
- **Laser Lights:** Laser illumination, also called structured illumination is used usually in microscopy applications, where a high precision illumination is required, or in applications where it is required to measure 3D features of the objects with accuracy. They provide a precise source of light applying different illumination patterns such as lines, points, crosses, circles and others. The use of patterns allow to measure different characteristics of the object, for example, a typical application is to use an oblique source of light and a line pattern in such a way that variations in the shape of the line will reflect small variations in the height of the object.

Fluorescent, quartz-halogen, and *LED* are by far the most widely used lighting types in Computer Vision, particularly for small to medium scale applications,

whereas metal halide, xenon, and high pressure sodium are more typically used in large scale applications, or in areas requiring a very bright source.

Figure 22 shows the advantages and disadvantages of fluorescent, quartz halogen, and *LED* lighting types. For example, whereas *LED* lighting has a longer life expectancy, quartz halogen lighting may be the choice for a particular application because it offers greater intensity.

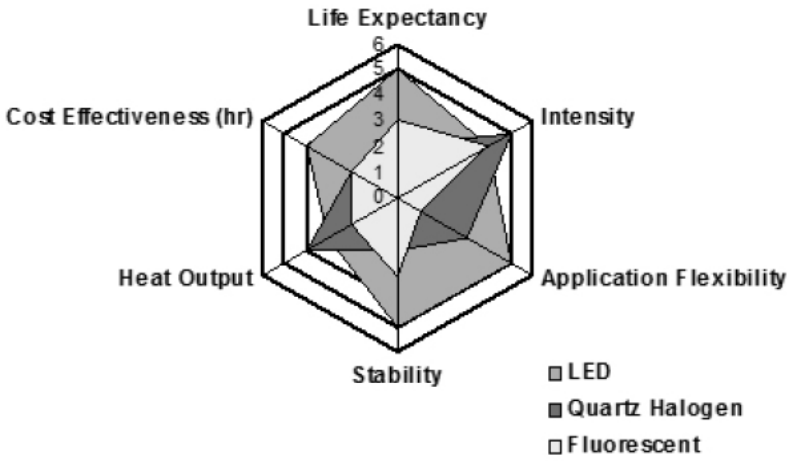

Figure 22: Comparison of most common light sources [81].

The most common lightning sources used in Computer Vision can be classified as follows:

Frontal illumination

This is the most common illumination technique. Specular light is used placing the camera looking to the object in the same direction as the light source. This type of illumination is useful in mate surfaces.

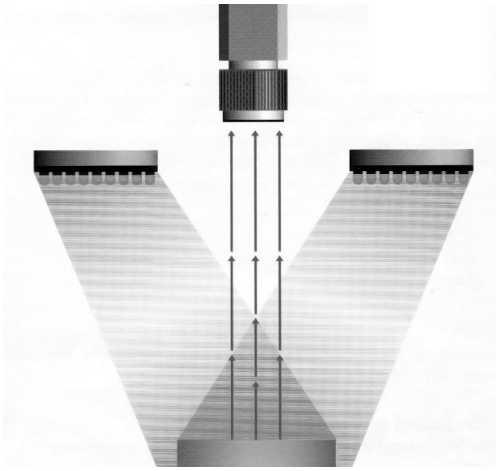

Figure 23: Frontal illumination [81].

Lateral illumination

Similar to frontal illumination, specular light is used placing the light source in angle; this can be useful to highlight certain characteristics of the object.

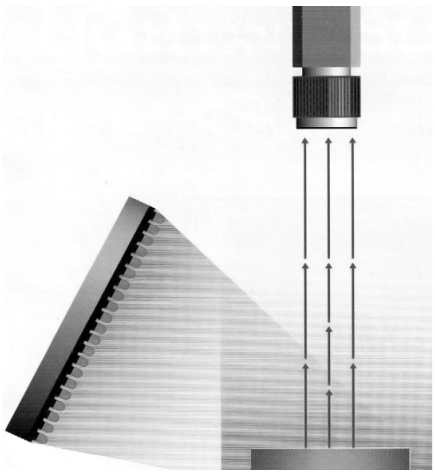

Figure 24: Lateral illumination [81].

Dark field illumination

Dark field illumination systems use diffuse reflection to illuminate the object by placing the light source in a low angle outside the bright field as shown in Figure 25.

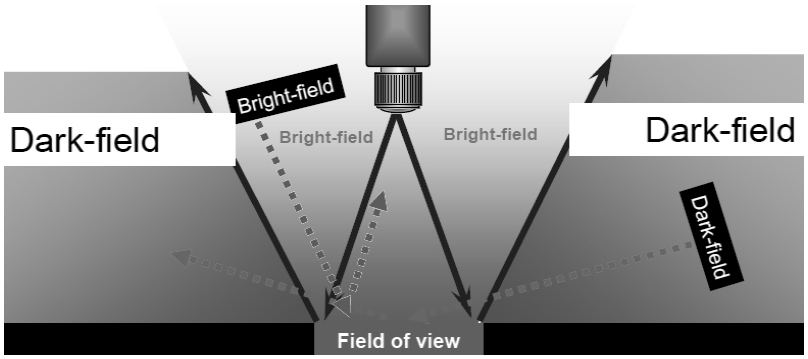

Figure 25: Bright field and dark field [82].

Dark field illuminators provide effective low-angle lighting to targeted regions. These lights enhance the contrast of surface features such as such as laser embossed, engraved marks or surface defects.

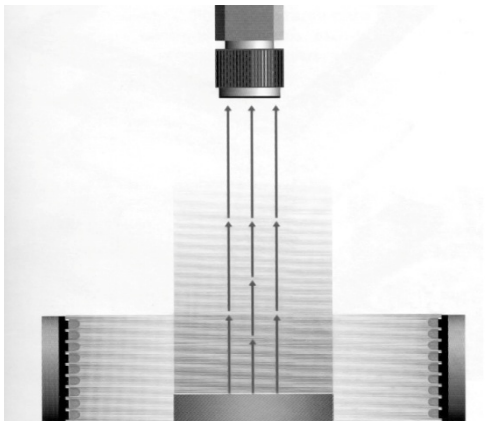

Figure 26: Dark field illumination [81].

Backlight illumination

In this illumination technique, light source is placed behind the object of interest pointing to the camera. Backlight illuminators provide sharp contrast to outline a part's shape, find edges and view openings such as drilled holes. However, this technique is not adequate to perceive the details of the surface.

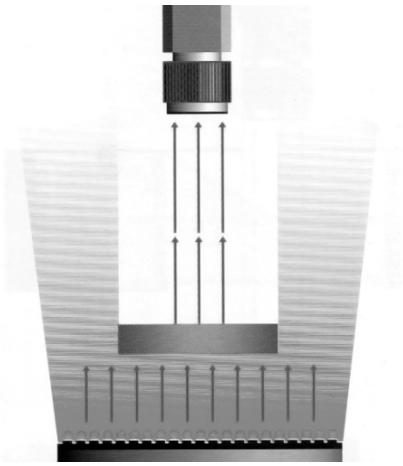

Figure 27: Backlight illumination [81].

Diffuse on Axis Light (*DOAL*)

With the *DOAL*, light rays reflect off a beamsplitter directly on to the object at nearly 90°. With this approach, specular surfaces perpendicular to the camera appear illuminated, while surfaces at an angle to the camera appear dark. Non-specular surfaces absorb light and appear dark.

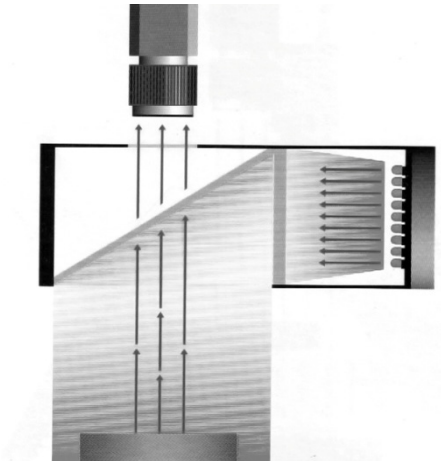

Figure 28: DOAL illumination [81].

Square Continuous Illumination (*SCDI*)

The *SCDI* works on the same principles as the *DOAL*, but with added uniformity for non-planar surfaces. With the *SCDI*, light rays reflect off the beamsplitter and the lower chamber, increasing the solid angle of illumination. The light source is tilted parallel to the beamsplitter increasing uniformity.

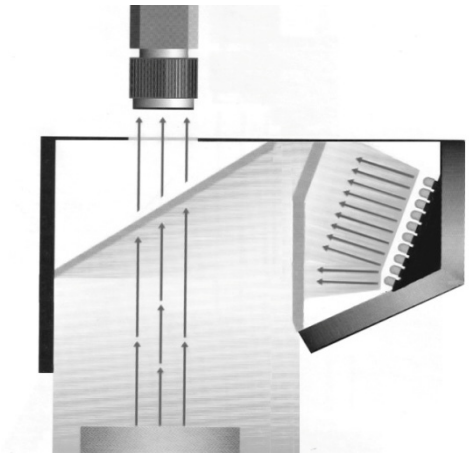

Figure 29: *SCDI* illumination.

Diffuse Dome Illumination

This technique illuminates the object with diffuse light reflected in a spherical dome. Dome illuminators are an economical source of diffused, uniform light. Their large, solid angle of illumination supports imaging of curved, shiny or bumpy surfaces.

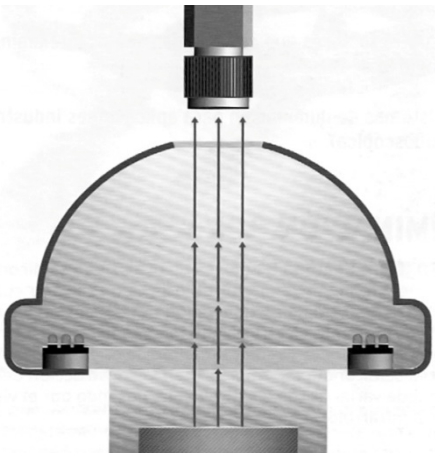

Figure 30: Dome illumination [81].

Cloudy day Illumination (*CDI*)

The *CDI* is based on the dome illumination, adding a second light source. It is ideal for the most complicated uneven and specular surfaces, because it offers the greatest degree of light coverage, nearly 170°, reflecting the light in a hemispheric pattern as wide as possible.

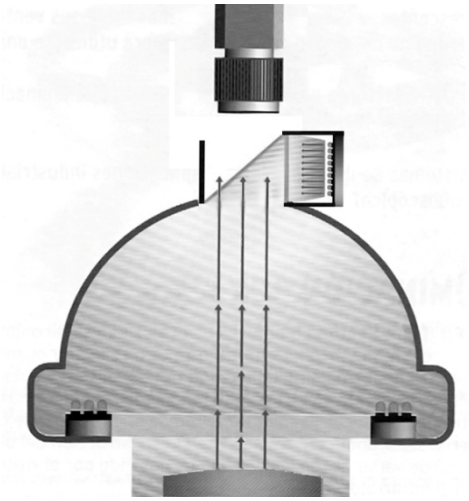

Figure 31: *CDI* illumination.

5.2.4 Illumination and Objects

One of the most important things in order to study the illumination of a scene is to compare the effects of various lightning techniques on part features of the objects of interest.

This is performed using a technique called *ATEST*, which purpose is to enhance or to mitigate different parts of the scene according to the interaction of light with different characteristics of the objects, creating the highest possible contrast between the feature of interest and its surroundings. The main object features in relation with light are summarized in Table 2 and can be described as follows:

Table 2: *ATEST* technique [83].

Part feature	Backlight	*DOAL*	Array, Ring	Dome, *SCDI,CDI*	Dark Field	Structured
Absorption: Look for change in light absorption, transmission or reflection	None	Uniformity of technique ensures absorption changes on FLAT surfaces are observable	Application dependent	Uniformity of technique ensures absorption changes on BUMPY surfaces are observable	Minimal effect	None
Texture: Look for change un surface texture or finish	None	Textured surfaces DARKER than polished	Application dependent	Minimizes texture	Textured surfaces BRIGHTER than polished	Some effect
Elevation: Look for change in height from surface to camera (z axis)	None	Angled surfaces are darker	Application dependent	Minimizes shadows	Outer edges are bright	Shows elevation changes
Shape: Look for change on shape or contour along x/y axis	Shows outside contours	Changes evident if background is different	None	None	Contours highlighted, flat surfaces darker than raised	None
Translucency: Look for change in density-related light transmission	Shows changes in translucency vs. opaqueness	Minimizes clear, FLAT overcoats (such as varnishes, glass), shows changes in translucency vs. opaqueness if background is different	Application dependent	Minimizes clear BUMPY overcoats (such as plastic over wrap, curved glass), shows changes in translucency vs. opaqueness if background is different	None	None

Absorption

The absorption of light makes an object dark or opaque to the wavelengths or colors of the incoming wave.

Some materials are opaque to some wavelengths of light, but transparent to others, for example glass and water are opaque to ultraviolet light, but transparent to visible light.

This property can also be used in the opposite direction, for example infrared light is effective at neutralizing contrast differences based on color.

Another manner that the absorption of light is apparent is by their color. If a material or matter absorbs light of certain wavelengths or colors of the spectrum, an observer will not see these colors in the reflected light.

Color of light can be used to take advantage of the absorption properties of the materials as shows Figure 32. Therefore, using the opposite light spectrum will make a part feature appear darker and using the same light spectrum will make a part feature appear lighter.

Figure 32: Absorption and light.
Can illuminated with red light (a). Can illuminated with blue light (b) [83].

The most simple, but also one of the most effective techniques to create a contrast between the feature of interest and the background in a controlled environment it is to modify the color of the background, for example painting the background with a color opposite to the color of the object of interest to highlight the shape of the object.

Texture

Texture, represent the visual quality of a surface, texture is affected by the direction of the illumination. For example, using Diffuse On Axis Light, polished

surfaces appear bright because they do not scatter and return light to camera while rough surfaces scatter light (see Figure 33).

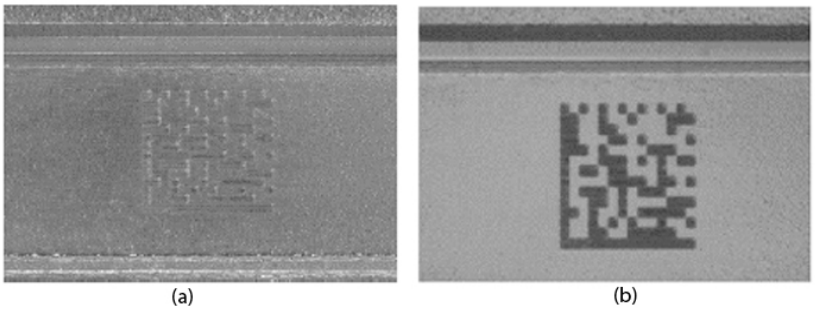

Figure 33: Texture and light.

(a) Data matrix illuminated with frontal light. (b) Data matrix illuminated with *DOAL* [83].

Elevation

The presence of 3D features such as elevations will cause light scattering. Therefore, the appearance of the object changes according to the direction of the illumination. For example, using a dark field illumination outer edges appear bright and using bright field illumination angled surfaces appear dark and flat surfaces appear bright (Figure 34) or using dome illumination elevations will not project shadows.

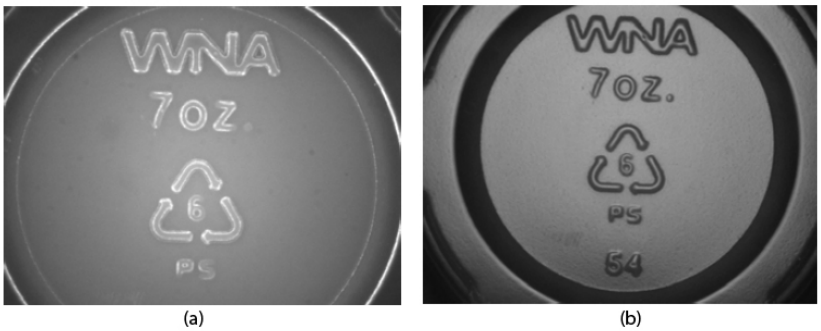

Figure 34: Elevation and light 1.

(a) Can illuminated with dark field light. (b) Can illuminated with bright field light [84].

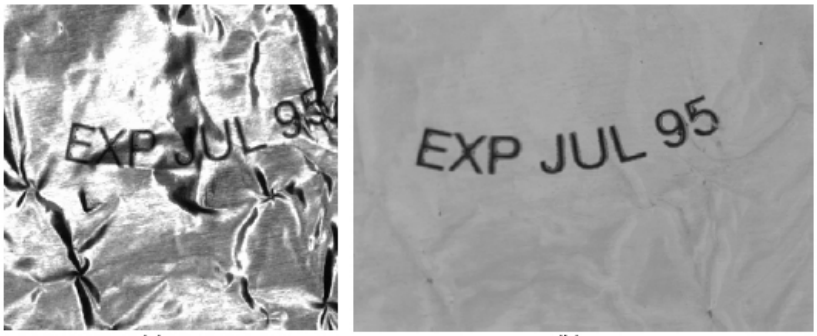

Figure 35: Elevation and light 2.

(a) Wrinkled foil with frontal light. (b) Wrinkled foil with Dome Illumination [83].

Shape

The perception of shape is also altered by the direction of light. For example outside contours can be highlighted using back light as shows Figure 36.

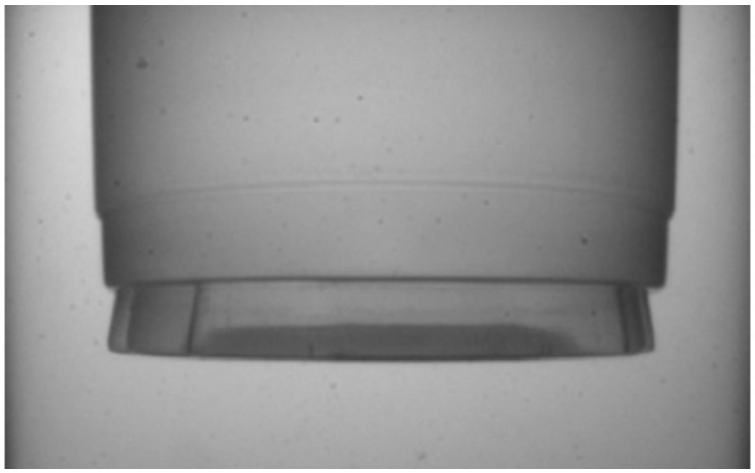

Figure 36: Shape and light.

Recipient illuminated using backlight illumination [84].

Illumination and Translucency

Is the physical property of allowing light to pass through a material, translucency is also affected by the direction of light, for example this property of the material is highlighted by using backlight (see Figure 37).

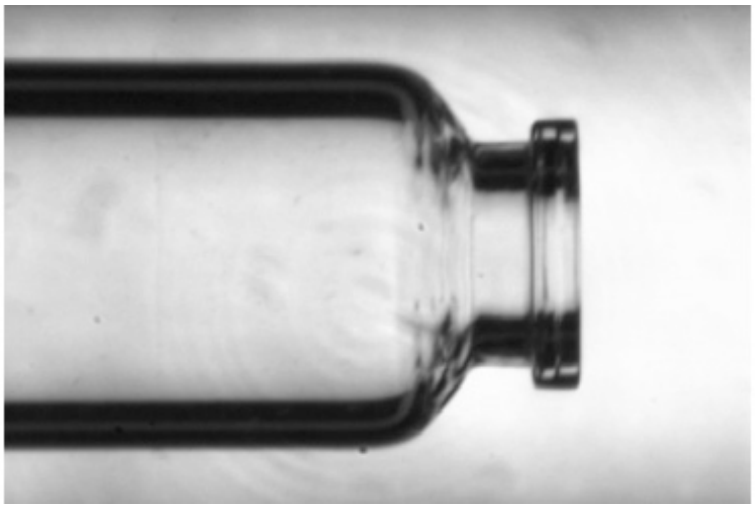

Figure 37: Translucency and light.
Glass bottle illuminated using backlight illumination [84].

5.2.5 Filters

A filter acts by absorbing a part of the light with certain characteristics, modifying the images recorded. There are several types of filters, being the following the most relevant for Computer Vision:

Figure 38: Optical filters [85].

- **High-pass Filter:** They transmit light above certain wavelengths.
- **Low-pass Filter:** They transmit light below certain wavelengths.
- **Band-pass Filters:** They transmit light within specific range of wavelengths they are used with structured laser illumination systems, to adjust the camera to the specific wavelengths of the laser.
- **Cut Filters:** They block specific wavelengths. For example the infrared or UV components of light, or a specific color of the visible light.
- **Contrast Enhancement:** They are colored filters commonly used in black and white imaging to alter the effect of different colors in the scene changing the recorded contrast for those colors.
- **Natural Density:** Is a filter of uniform density which attenuates light of all colors equally.
- **Polarizer:** Light reflected from a non-metallic surface becomes polarized. That is, the electric field of the wave acquires a direction; this effect is maximum at about 56° from the vertical.

  Polarizing block light polarized in a specific direction. Polarizing filters are usually applied in pairs, Polarizing filters, one between the light and sample

and the other between the sample and camera, typically affixed to the lens via screw threads,.

They are useful to reducing glare or to detecting differences in mechanical damage in otherwise transparent samples (see Figure 39).

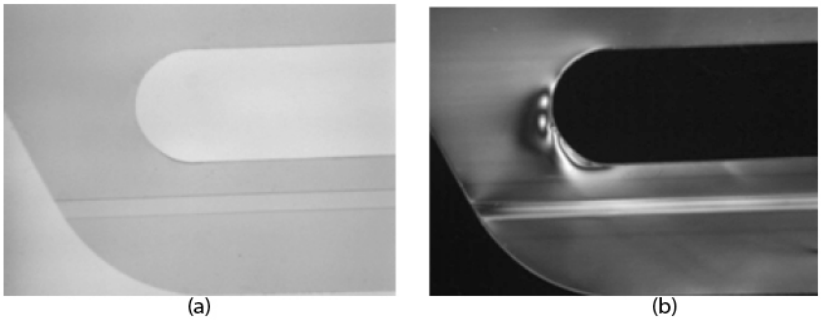

Figure 39: Detecting mechanical damages with filters.

Transparent plastic 6-pack can holder illuminated with (a) red back light (b) Red back light with a polarizer pair, showing stress fields in the polymer. [86].

Every master knows that the material teaches the artist
Ilyá Grigórievich

V. Application in Material Testing

One of the main issues in Civil Engineering is to analyse the behavior of materials in strength tests. Traditionally, information about displacements and strains in the materials is carried out from these tests using physical devices such as strain gauges or other transducers.

Although these devices provide accurate and robust measurements in a wide range of situations, they are limited to obtaining measurements in a single point of the material and in a single direction of the movement, thus, the global behavior of the material cannot be analysed.

On the other hand, in Computer Vision, the measurement of non-rigid movements is one of the most studied topics. In particular, due to their accuracy and

robustness, Block-Matching techniques have been widely used in the task of estimating the displacement field in a scene.

In this chapter, Block-Matching algorithm is presented to analyses the behavior in 2D deformable surfaces, and in particular in the analysis of strength tests in Civil Engineering.

The proposed technique consists of the integration of a calibration process with a new Block-Matching technique.

The Block-Matching algorithm uses an iterative analysis with successive deformation, search, fitting, filtering and interpolation stages. It uses previous displacement fields to compute deformations in long sequences and it can use different metrics, sub-pixel algorithms and warping techniques.

The proposed algorithm has been designed to measure displacements from real images of material surfaces taken during strength tests, and it has the advantages of being robust with long range displacements, of performing the analysis on the image space domain (avoiding the problems related to the use of Fourier Transforms) and of measurements in long sequences related to a real point of the material and not to a pixel position.

To validate the proposed approach different experiments were carried out. First, a set of image sequences was analysed to evaluate performance and accuracy of the technique, using different metrics, subpixel methods and deformation models to provide for the first time, a study of the influence of these parameters in the accuracy of a Block-Matching technique.

Then, a comparative with different state of the art optical flow techniques has been performed, providing a context with different techniques and motions.

Finally, different test were performed to evaluate the performance of the algorithm in the field of strength tests, both in simulated and in real conditions.

1 Introduction

Some of the main needs in Civil Engineering are to know the stress-strain response of materials used structures, to evaluate their strength and to determine their behavior during the design working life.

For this purpose, strength tests are usually carried out by applying controlled loads or strains to a test model (a scale model of reduced size, maintaining all the properties of the material to be analysed). The goal is to find out parameters such as material's strength or displacement evolution, comparing real construction models to theoretical ones or determining whether a given material is suitable for a particular use.

In these tests, information about the material's behavior is traditionally obtained using specific devices, such as strain gauges or *LVDT* sensors, which are physically linked to the material and provide information about the length variation of the structure around a given point and in a particular direction.

These devices are limited, in the sense that they only provide readings in one dimension and with a given accuracy, which is limited by their construction and by the device functioning. Moreover, they generate certain stiffness to displacement, thus interfering with the experiment and some of them, such as strain gauges, cannot be reused.

Figure 40 shows the way traditional sensors are used.

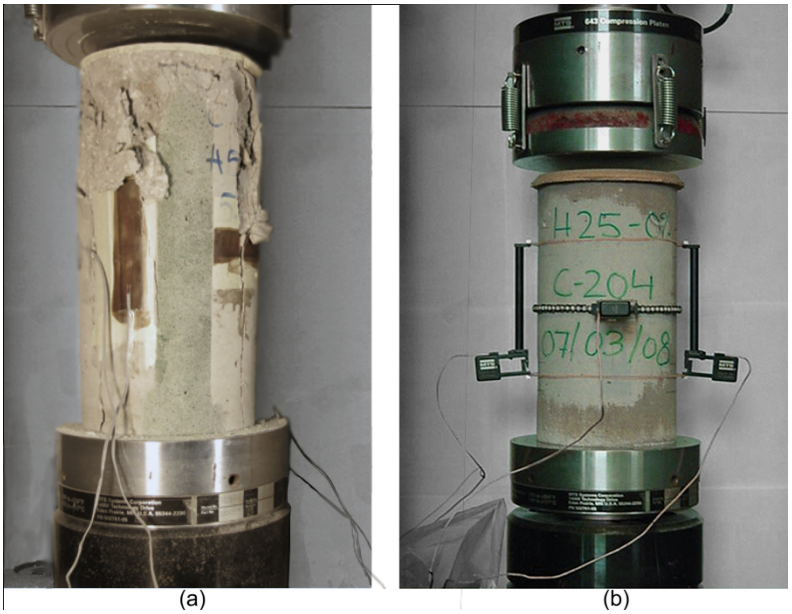

Figure 40: Traditional sensors for strength tests.

Traditional displacement measurements in strength test. Sensors are attached to the specimen and each sensor provides a direction reading in a single point. (a) Shows strain gauges in a concrete model after a compression test. (b) Shows contact extensometers in a similar model.

Finally, although in recent years some non-contacting systems such as mechanically driven optical followers, laser extensometers and video extensometers are used, these devices still operate on a single measuring line or they need to mark the measurement points before the test.

The proposed technique uses a video camera to record the test and then processing the images to extract the displacement field of the body, that is, a vector field of the displacement for all the visible points on the surface of the body.

The theoretical advantages of this process compared to traditional systems are as follows:

- The measurement does not interfere with the experiment.

- Displacement is measured in multiple points, in multiple directions at the same time, so that the global displacement field is obtained.
- System has a low cost and can be reused.

The main contribution of this research is the study of strength tests using Block-Matching techniques to calculate the strain and the displacement field in the material.

To this end, a Block-Matching algorithm has been created. This technique quantifies the deformations in the media, using a new procedure to analyse image sequences and it uses calibration and image correction techniques to obtain real scale measurements.

The main characteristics of the proposed Block-Matching algorithm are the following:

- It uses a multi-resolution one by using an optional pyramidal decomposition stage.
- The model was uses an iterative prediction correction process integrated with several deformation models, (including a new local parametric deformation model).
- The algorithm provides a dense vector field output, obtaining a measurement for each pixel in the image, instead of obtaining a single measurement for each block.
- The proposed model supports several similarity metrics.
- The proposed model supports several subpixel techniques.
- A new technique to search in image borders called Floating Block Technique, has been implemented in the algorithm.

- A new technique to perform measurements in long image sequence, controlling the increment from adding partial results has been implemented in the algorithm.

Using the proposed technique, a comparative with different metrics, sub-pixel algorithms and deformation models was performed in standard synthetic and real image sequences. Some of these techniques may be used for the first time with Block-Matching and in most cases their application to Block-Matching may have not been properly addressed before. The obtained results were compared as well with different state of the art algorithms, to provide an appropriate context.

The proposed technique was used in a real scenario, to solve the medical problem of matching proteins in images of 2D electrophoresis gels, improving previous published results.

2 State of the Art

In the challenge of finding a new way of analysing the behavior of materials in strength tests, *CV* techniques can obtain information very different from the traditional sensors, thus a simple image contains much more information of the world than traditional one parameter lectures from a single point.

CV techniques were applied for the first time in industrial applications in the 80's. [87]. Nowadays, they are successfully used in several fields such as object recognition [88], robotics and navigation [89] or, more related with this research, surface characterization of materials [90, 91] and tracking of moving objects [92, 93].

Recently, some *CV* applications have been proposed in the analysis of strength tests. The most important contribution is the video extensometer. A video extensometer is a device to perform non-contacting measurements by placing special marks on the specimen and tracking these marks with *CV* techniques. With these techniques, changes in the distance (longitudinal strain) between a pair of marks can be measured. These devices have been studied in different works [69, 70] and have been commercialized [94].

However, although the video extensometer is a mature technology, it can only perform measurements in a small set of predefined points (so the main effects of deformation can be missed), and don't allow the full characterization of the material.

The main difficulty for analysing displacement in non-rigid bodies from visual information is derived from the fact that the body geometry varies between two instants and thus, it is not possible to find out which part of the information characterises the body and can be used to estimate the displacement.

Traditional techniques for analysing strain processes in Computer Vision are based on finding the parameters of a displacement model. Some additional knowledge about the model must be added to these procedures in order to obtain some robust measurements, considering non-rigid movements. This knowledge typically consists of a set of well-known correspondence points or estimating the equations regulating displacement $d(i, j)$ [95, 96].

Optical flow techniques were carried out by Horn and Schunck [51] , and these are a field of Computer Vision for displacement analysis. These techniques provide a flexible approach regarding the extraction of the motion field of a scene without using any previous knowledge about the displacement of the objects in the image.

Nevertheless, only very limited or too costly applications of these techniques have been found so far in the field of Civil Engineering [97].

The works carried out in the flow-field analysis [98, 99] showed that when the analysed scene does not contain multiple objects with different motions, a Block-Matching technique is the most robust and flexible approach.

Region analysis or Block-Matching techniques calculate the displacement of a point *(i, j)* in the image *I*, analysing the statistical similarity of the image in its surroundings, with the region in image *I′* centered at *(i′, j′)* and representing a possible displacement of the original region, as shown in Figure 41:.

Along time, several works have been focused in Block-Matching techniques and important contributions have been carried out.

In order to increase accuracy, sub-pixel estimation using parabola fitting over three points was used in several works [98, 100].

Also, multi-resolution approaches to decrease computational cost and to avoid local minimums have been widely proposed [101-103]. A related contribution is the adaptative reduction of the search area [104, 105].

Using a different perspective, feature selection techniques had been included in some works [35] to select optimal initial blocks.

Other important contribution was the analysis of displacements in the frequency domain using Fast Fourier Transforms (*FFTs*) to increase performance [98, 106].

Recently, some new methods have been proposed to explicitly include deformation in the search process. Essentially, in these works optical flow is calculated in an iterative prediction-correction process [99].

Based on early iterative image registration principles [50], some new methods have been proposed to explicitly include deformation in the search process [99].

In these techniques, dense fields are calculated by interpolating the block displacement field. Then, a second interpolation process is used to simulate deformation on the image. Therefore, in the next iteration the results are improved performing a new analysis with the deformed image.

Additionally some strategies have been included in order to compensate the process instability, such as corrector filtering [107].

Some examples of iterative techniques are based on the use of B-spline interpolation [108], first-order bidimensional Taylor series [99], or Radial Basis Functions (*RBF*) [101, 109] to interpolate the vector field of displacement.

In addition, some works have used different approaches to improve the obtained flow, some examples are the proposal of a statistical approach [110, 111], the use of Particle Swarm Optimization (*PSO*) technique [36, 112], the inclusion of edge analysis [113], the use of feature information [114, 115], or the use of Markov Random Fields (*MRF*) [116].

Other minor contributions have included the study of new similarity measurements [117] or new search strategies [118-120].

Currently, some works are using GPUs [44] and hardware [42] architectures to improve efficiency of Block-Matching techniques, while others are introducing new techniques such as Kalman filters [95], evolutionary computation [45] and even artificial life [48].

The main advantages of block-matching techniques are simplicity, flexibility, robustness and locality. However, these techniques are limited due to the block concept itself. Because a block has a size and shape defined *a priori* and which lead to wrong measurements near discontinuities in the motion field [121, 122].

Several papers have attempted with some success to reduce this problem by using a multi-layer approach [123], pairwise affinities based on boundaries [121] adaptive shape windows [124-129], adaptive support-weight windows [130], barycentric correction [131], by feature matching methods [132] or stereo analysis of cost volumes [122]. But the question is still far from a standard solution.

Currently, it can be assumed that to analyse a scene with different objects moving with different motions, a Block-Matching technique may not be the best approach when the accuracy is the most important factor.

Therefore, Block-Matching has been applied in several fields such as in the analysis of flows [99], the analysis of deformable materials [49], video processing [133] or medical image analysis [43, 87, 95, 134].

Currently, Block-Matching is one of the most robust methods for extracting the displacement field of a surface without reference points such as corners or edges.

3 Formulation of the Problem

Measuring displacement in deformable media is an issue with a great potential application to various fields, ranging from bioinformatics to Civil Engineering.

To estimate the displacement in a sequence of images, most techniques use a fundamental assumption called the brightness constancy assumption (10):

$$I(x,y,t) \approx I(x+\delta x, y+\delta y, t+\delta dt) \quad (10)$$

Where I is the intensity value of a pixel in a defined position and concrete time. And $(\delta x,\ \delta y)$ represent the displacement of the point located at (x, y) in the time interval δdt , which can also be expressed as in (11):

$$(\delta x, \delta y) = d(x,y) = (dx, dy) \quad (11)$$

From the point of view of image processing, the previous assumption can be defined in a more practical way, considering two states (or images) I and I' which represent the same scene in different moments. Therefore, movement can be considered as a system whose input I generates an output I' determined by system's transfer function, which is integrated by the displacement function $d(x, y)$ plus a noise signal N_S. This assumption, which is analogue to the previous one, is expressed in (12).

$$I(x,y) + N_S = I'(x+dx, y+dy) \quad (12)$$

To solve the previous system and find out the displacement, it is necessary to include further restrictions in the model. And this is usually done by using the neighborhood of the selected pixel.

Block-Matching techniques calculate the displacement of points (x_i, y_i) defined in I by comparing the region or block around this pixel with candidate blocks defined in I'. Representing each of blocks possible displacements of the original region.

This is based on the assumption that displacement is constant in each block, i.e. that each block suffer a linear displacement.

Therefore, the block in *I'* centered at the point (x_i', y_i') and obtaining the largest similarity value with the original one, will be selected as the solution. The general Block-Matching procedure is shown in Figure 41.

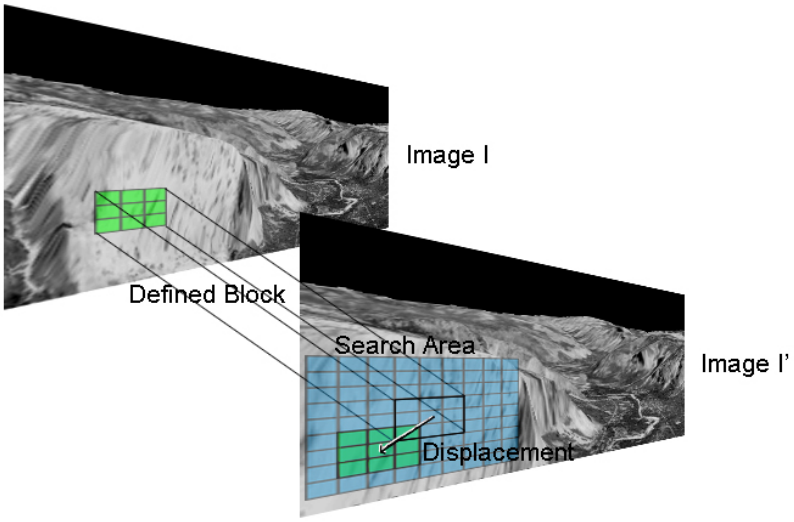

Figure 41: Block-Matching technique.

One block in a frame is compared with multiple regions representing different possible displacements in the next frame

4 Block-Matching with Non-rigid Displacements

When non-rigid movements are considered, the challenge consists in extracting the displacement in a sequence where the geometry of the bodies varies between two moments.

The problem lies in the fact that traditional Block-Matching model considers only linear displacements (first order effects), without taking on board rotation or

deformation (second order effects) which characterize deformation movements. Although modern techniques include more complex models, it is still necessary to perform a linear estimation in the first stages of these algorithms.

To this end, Block-Matching techniques make the assumption that, if the regions or blocks are small enough, then the second order effects may be disregarded (in the last instance, an infinitesimal point will experience a linear displacement); so body deformations may be calculated from the locally linear displacements in every region.

Thus, the block size controls the deformation rigidity which can be measured, so that the bigger is the region, the lesser is the flexibility in the measurement of deformation.

It should also be considered that by decreasing the time interval between the compared images, the object displacement and deformation effects will also decrease [135].

In practice, this leads to the next condition: Time interval and block size should be small enough to assume that the displacement can be estimated using a linear approximation of the displacement. This implies that the shape of the block will not suffer strong variations after the displacement as it is expressed in (13).

$$B_{xy}(p) \approx B'_{x'y'}(p) \quad (13)$$

Where B_{xy} and $B'_{x'y'}$ represent two blocks centered in *I(x, y)* and in *I'(x' ,y')* respectively. These blocks can be represented as ordered vectors (or matrixes) of intensity values in the neighborhood of the block center. The value *p* expresses the position of an intensity value inside the block. In this context, the displacement of the pixel *(x, y)* is *d(x, y) = (u, v)* being *(x+u, y+v) = (x', y')*.

This assumption is in practice, both a brightness constancy assumption and an assumption about the softness of the deformation inside the block.

However, the presented locality condition have two implications that most Block-Matching techniques do not face properly; first, the use of several pyramidal decomposition stages may lead to an infringement of this condition (since image size is reduced and therefore, the block size is augmented). And last, measuring the total displacement in a sequence from the initial frame may entail another infringement of the locality principle (since the time interval grows with each frame).

The presented Block-Matching technique uses an iterative process to calculate displacements. A linear estimation of the movement is obtained in the first iteration, and deformations are computed in next iterations using previous estimations.

Additionally, the locally linear assumption of displacement is only used between consecutive frames to calculate total displacements from a sequence, and the pyramidal stage is optional. Therefore, the presented technique can be used to analyze complex deformation processes occurring along time.

5 Proposed Technique

5.1 General Functioning

The proposed algorithm (summarized in Figure 42) calculates real scale motion vectors from images of a recorded test.

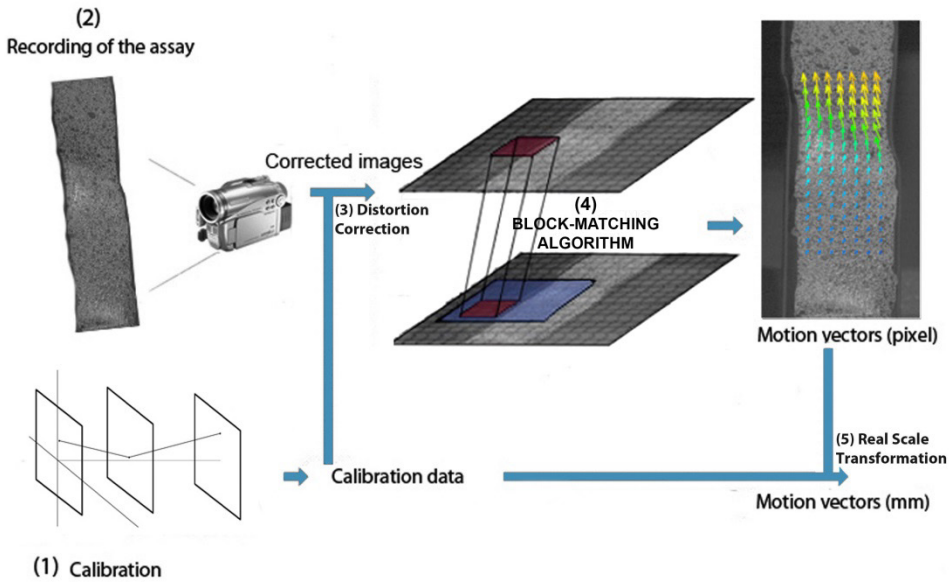

Figure 42: Proposed technique.

Outline of the proposed technique from images of the test to motion vectors on a real scale.

The main processing steps are the following:

1. The camera is calibrated. For this purpose, the parameters and the position of the camera must be selected and set. They will remain constant throughout the test.
2. The test is performed and recorded.
3. The distortion of the test images is eliminated using the information from the lens distortion model and the calibration data.
4. Motion vectors are calculated in the corrected images using the Block-Matching technique.
5. Calibration data are used to transform motion vectors from pixel to real scale.

The Block-Matching algorithm divides the image regions called blocks and solves the correspondence problem for each block in the next image. To this end, it uses

an iterative approach with different stages. The result is the most likely displacement for each original region. A general scheme of the algorithm is shown in Figure 43.

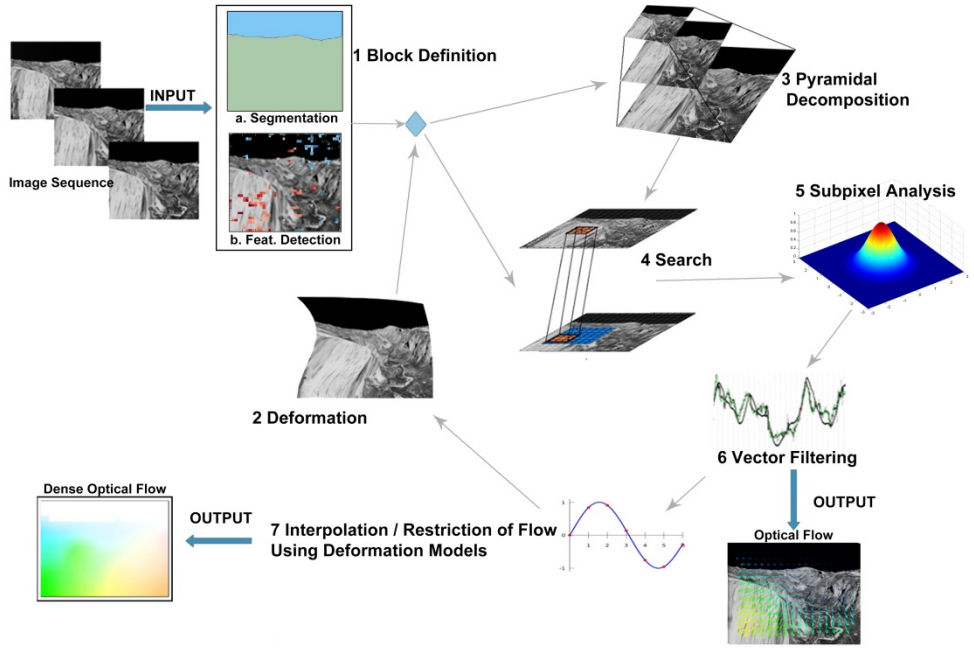

Figure 43: General scheme of the proposed Block-Matching algorithm.

The main steps of the algorithm to analyse the displacement between two frames (a reference and an objective one) are the following:

1. Block Definition: Blocks are defined taking in account the shape of the different objects (calculated with a segmentation technique) and around points of interest (obtained with a feature detection technique).

2. Deformation: The search area and the shape of the target blocks are updated using the displacement field from the previous iteration. To obtain the new blocks, the previous deformation model is used with a bilinear interpolation process on the image.

3. Pyramidal decomposition: This technique consists in processing the image using a hierarchy of scales. To this end the image is iteratively downsampled while the block size remains constant and the search range is reduced. This technique is useful to avoid local minimums and to reduce the computational time. The number of pyramids can be selected by user.
4. A similarity metric and a search strategy are used to calculate the best displacement for each block.
5. The similarity values are used to perform a fitting to a selected function with a selected fitting technique to achieve a higher accuracy
6. The obtained vectors are filtered and smoothed.
7. The Optical Flow is processed with a deformation model used to interpolate the displacements and to impose restrictions to the displacement field. In the next iteration, an estimation of the deformed image will be obtained from the dense field as indicated in in point 2).

To compute the total displacement in a sequence, an addition of partial displacements is performed. To this end the source frame and the source block are dynamically updated using the previous displacements.

With this procedure, the source block is updated and it follows the position of the original point through time, so that the addition of displacements has a physical meaning: the total displacement experienced by the object at a given point of its surface.

In this scenario, the error accumulation due to the addition of partial results must be considered. The influence of the error factor in a displacement measurement can be expressed as follows:

$$D_R^i + E^i = D_M^i \quad (14)$$

Where D_R^i is the instant displacement of the point at the moment i, D_M^i is the measured displacement and E^i is the error obtained from the measurement. Thus, when the total result comes from a sufficiently high number of additions, the influence of the error factor will grow unpredictably as showed in (15):

$$D_R^T + \sum_i E^i = \sum_i D_M^i \quad (15)$$

In order to mitigate the effects of error accumulation, an approach which monitors the number of frames to be compared with each base image has been chosen, changing the image used as source after a certain number of images, or when the average displacement reaches a selected threshold. Thus, the total error influence would be expressed as follows:

$$E^T = \frac{\sum_i E^i}{N} \quad (16)$$

Where N is the number of frames to be compared with the same source frame.

Taking this into account, displacements from previous frames are used to update the search area and the shape of the block, even if source image and source blocks are not updated. This is performed in the same way as in previous iterations to estimate deformation for the next ones for the same frame.

With this approximation, the total displacement can be calculated without increasing too much the time interval between the frames and without violating the locality condition.

5.2 Calibration Process

In order to obtain real scale information of the measurements in the material, the pin-hole camera model has been used as described by Zhang [27] with the distortion model proposed by Weng et al. [74].

The model was solved with the standard calibration method proposed originally by Zhang [27], which is currently the most used calibration method. This technique was described with detail in section IV.3.3.

5.3 Segmentation

As expressed before, one of the most important limitations of Block-Matching techniques is the presence of discontinuities in the motion field caused by objects with different motions.

Although different techniques may be proposed to solve this limitation, the problem can only be solved by introducing knowledge about the objects of the scene (or their edges).

The solution proposed in this paper is one of the most general approaches, consisting in using a segmentation technique to define the blocks of interest.

Segmentation is one of the main issues in computer vision, and can be defined as the problem of partitioning the image into non-overlapping regions, representing each region a different object in a scene.

Several segmentation algorithms have been published so far, ranging from K-means clustering [7], Background subtraction techniques [8], active contours [6],

region growing techniques [42], Artificial Neural Networks [136] and many others.

The presented Block-Matching technique can use the information from any segmentation algorithm to modify the shape of the blocks defined in the source image. Furthermore, every processing stage of the algorithm operates independently for each detected object, so different deformation models are applied for different objects and the filtering stage is based on neighborhood relations in the same object.

In this paper, since it is necessary to isolate the segmentation error from the error due to the Block-Matching algorithm, a semiautomatic technique has been selected. It is based on use of a gradient descending technique to perform an approximation of a preliminary user segmentation to real boundaries. This is a technique frequently used in biomedical applications [9].

5.4 Feature Detection

Once separated the objects in the image, it is necessary to define the different blocks which will be used to perform the matching process. Typical Block-Matching techniques cover the whole image using a two dimensional regular grid, allowing the use of some fast computation techniques such as kernel filtering and bilinear interpolation.

In the proposed solution any distribution is possible with the only limitation of requiring a minimum number and density of points in an object to obtain valid results when using some deformation models and to make the filtering stage work properly.

This allows the user to choose from very different techniques to define the desired blocks; ranging from marking manually the image to using a normal distribution or using an automatic technique to detect points of interest

5.5 Similarity Metric

The algorithm proposed in this research measures movements using the statistical similarity of the grey levels in each region of the image. For this purpose, a region or block is formally defined as a sub-area of the image of a particular and constant size and shape.

At this point, a similarity metrics, robust to natural variability processes in real images are needed to compute the most probable displacement for each point. This is important due to the natural changes in refraction in moving objects, changes in visual properties due to deformations, and the variability present in common light sources.

Therefore, given a block B_{xy} the similarity value $S_{u,v}$ will be the result of comparing the initial block with a candidate one $B'_{x'y'}$ using a similarity function $f(B_{xy}, B'_{x'y'})$ which output measures the confidence in the assumption expressed in (12) and being $d(x, y)=(u, v)$ a displacement accomplishing $(x+u, y+x)=(x', y')$.

In the proposed technique, different similarity metrics have been used, and they can be selected to solve problems with different requirements.

Mean Squared Error (MSE):

The mean squared error *(MSE)* is a popular estimator to quantify the difference between two variables. So minimizing the *MSE* value is the same as maximizing a similarity metric.

A normalized version of *MSE* defined in (17) has been used in here.

$$MSE(B_{xy}, B'_{x'y'}) = \frac{\sum_{p}\left(B_{xy}(p) - B'_{x'y'}(p)\right)^2}{\sum_{p}\left(B_{x'y'}(p) - \mu\right)^2} \quad (17)$$

Pearson correlation quotient (R):

The Pearson correlation quotient measures the strength of linear dependence between two variables. It is defined as the covariance of the two variables divided by the product of their standard deviations (18).

$$R(B_{xy}, B'_{x'y'}) = \frac{\sum_{x,y}\left(\left(B_{xy}(p) - \mu\right) \times \left(B'_{x'y'}(p) - \mu'\right)\right)}{\sqrt{\sum_{p}\left(B_{xy}(p) - \mu\right)^2 \times \sum_{p}\left(B'_{x'y'}(p) - \mu'\right)^2}} \quad (18)$$

The Pearson correlation quotient has the advantages of being invariant to the average gray level and therefore it is robust in presence of some natural variability processes such as those present in common light sources.

Combined Similarity Metric (*CSM*):

Combination of different similarity techniques has been widely proposed [137, 138] where usually the registration parameters are weighted according to the strength of each measure.

In this research, a metric called Combined Similarity Metric *(CSM)*, formed by a combination of the Pearson's correlation quotient (*R)* and the Mean Squared Error (*MSE)* has been used. This measure was proposed in [139] and was used in [49] with Block-Matching techniques. It is defined as follows:

$$CSM(B_{xy}, B'_{x'y'}) = \frac{R(B_{xy}, B'_{x'y'})}{1 + k \times MSE(B_{xy}, B'_{x'y'})} \quad (19)$$

Where k is a constant for weighing the *MSE* in the measurement, so that $k = 0 \rightarrow CSM = R$.

This constant was empirically adjusted to $k = 0.15$ using a training databank granted by Otago University, New Zealand [140], where the proposed metric shown a better performance than *R* and *MSE*.

Mutual Information (*MI*):

Mutual Information of two variables *X*, *Y* can be expressed as the amount by which the uncertainty about *Y* decreases when *X* is given: the amount of information *X* contains about *Y*. and vice versa [141].

Mutual information equation is expressed in (20):

$$MI(X,Y) = H(X) + H(Y) - H(X,Y) \quad (20)$$

Where *H(A),H(B)* and *H(A,B)* are the entropy of *A*, the entropy of *B* and the joint entropy of *A* and *B*. which means that maximizing mutual information is related to minimizing joint entropy.

Entropy can be viewed as a measure of uncertainty. Here, the Shannon adapted measure was used [142] and as expressed in (21).

$$H(X) = -\sum_{x} p(x) \times \log p(x)$$
$$H(X,Y) = -\sum_{x,y} p(x,y) \times \log p(x,y) \quad (21)$$

Where *p(a)*, *p(b)* and *p(a, b)* are the probability of *a*, the probability of *b*, and the joint probability of *a* and *b* respectively.

The Mutual information can be obtained from Entropy. In this disertation, a normalized form of Mutual Information defined in (22) have been used, and distributions and joint distributions for the blocks B and B' have been estimated by means of histogramming [143].

$$MI(B_{xy}, B'_{x'y'}) = \frac{H(B_{xy}) + H(B'_{x'y'})}{H(B_{xy}, B'_{x'y'})} \quad (22)$$

While Mutual information requires a heavy computation time, it has the advantage over correlation methods of being capable of analyse multimodality images.

The measure has also been shown to be well suited to registration of images of the same modality [141].

5.6 Search of the Similarity Peak

The selected search function will be used to analyze the candidate blocks in the search area. In this point, it is possible to use different search strategies to reach the maximum similarity peak. Most of current techniques perform a search with discrete candidate displacements.

The discrete search process can be exhaustive, granting the obtaining of the maximum in the search space, or it can use only a few values of the search area trying to find the maximum using less computational time, some examples of these strategies are: the Three Step Search (*TSS*), the Diamond Search (*DS*) or the Hexagonal Search (*HS*) [119].

In this thesis, only the exhaustive discrete search has been used, although the proposed technique can be used with any search strategy.

The data obtained in the similarity analysis lead us to the most probable discrete displacement. However, the correlation value itself contains some useful information, given that the correlation values achieved in the pixels surrounding the best value will reflect a part of the displacement located between both pixels.

Thus, in order to achieve these measurements below the pixel level, the correlation values are translated to a continuous space using a fitting technique.

In this research, different functions and techniques have been used to perform this fitting.

5.6.1 Search in the Image Borders

One of the problems of Block-Matching techniques is the difficulties of searching on the borders of the image. This is due to the fact that an object can move towards the image border being a part of it outside the image limits, so correlation with the original block can't be performed with most of Block-Matching techniques.

In this research, a new technique called floating block is proposed. With this technique, a mask is used to simulate a displacement of the original block outside the image limits, so a fraction of the block size can be used for correlation. The minimum block size needed to perform a correlation can be selected by user. Figure 44 illustrates how the proposed technique works.

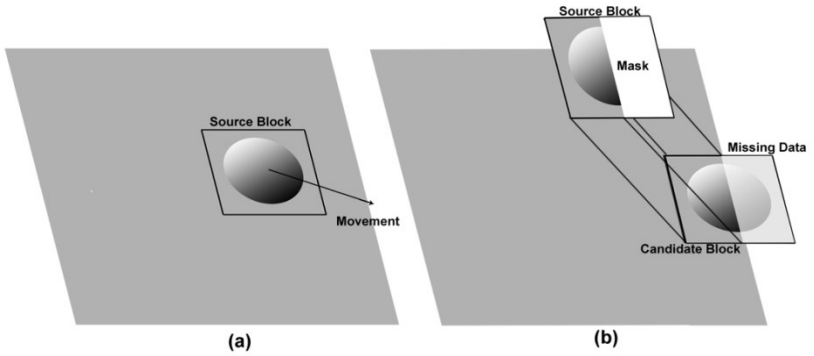

Figure 44: Floating block Technique.

(a) Initial frame: a source block is defined around a moving sphere. (b) The sphere has moved to the image border and a part of its shape is lost outside the image. The source block is compared with the current image using a mask which represents the missing information.

5.6.2 Three Point Estimators

One of the most used methods for estimating sub-pixel displacement is to find the true peak of the similarity function by fitting a parabola or other functions over three indexes near its limits [98-100]. This idea is illustrated in Figure 45.

This model uses separated one dimensional functions for each dimension of the movement and only three points are needed to solve each equation. Usually, one of the points is the maximum discrete peak and the other two are points at a symmetric distance of the maximum in the considered dimension.

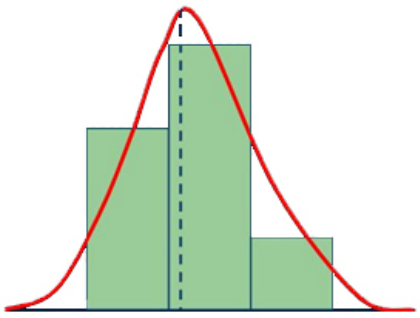

Figure 45: Proposed technique.

The advantages of this model are the simplicity and speed. However, this kind of estimators is subject to systematic errors which have been studied in different works [144].

In this thesis, three point estimators have been used with Gaussian (23) and polynomial (24) functions.

$$f(x) = \lambda \times e^{-\left(\frac{(x-\mu_x)^2}{2\times\sigma_x^2}\right)} \quad f(y) = \lambda \times e^{-\left(\frac{(y-\mu_y)^2}{2\times\sigma_y^2}\right)} \qquad (23)$$

$$f(x) = a \times x^2 + bx + c \quad f(y) = a \times y^2 + b \times y + c \qquad (24)$$

Using the previous functions, the corresponding three point estimators are defined in (25) and (26).

$$x_0 = x' + \frac{\ln\left(S_{u-1,v}\right) - \ln\left(S_{u+1,v}\right)}{2\times\ln\left(S_{u-1,v}\right) - 4\times\ln\left(S_{u,v}\right) + 2\times\ln\left(S_{u+1,v}\right)}$$
$$y_0 = y' + \frac{\ln\left(S_{u,v-1}\right) - \ln\left(S_{u,v+1}\right)}{2\times\ln\left(S_{u,v-1}\right) - 4\times\ln\left(S_{u,v}\right) + 2\times\ln\left(S_{u,v+1}\right)} \quad (25)$$

$$x_0 = x' + \frac{S_{u-1,v} - S_{u+1,v}}{2\times S_{u-1,v} - 4\times S_{u,v} + 2\times S_{u+1,v}}$$
$$y_0 = y' + \frac{S_{u,v-1} - S_{u,v+1}}{2\times S_{u,v-1} - 4\times S_{u,v} + 2\times S_{u,v+1}} \quad (26)$$

Where $S_{u,v}$ is the maximum similarity value for the block B_{xy} corresponding to the block $B'_{x'y'}$ and the displacement $d(x, y)=(u, v)$ where $(x+u, y+v)=(x', y')$. The values (x_0, y_0) are the new sub-pixel estimated values.

5.6.3 Levenberg-Marquardt Technique

A more complex, but also more accurate solution can be reached using a two dimensional function and a least squares fitting algorithm.

In this research, a variant of the Gauss-Newton (*G-N*) technique, called the Levenberg-Marquardt (*L-M*) technique has been used. An example of *L-M* fitting is illustrated in Figure 46.

Two dimensional Gaussian (27) and polynomial (28) functions have been used to model the fitting curve.

$$f(x,y) = \lambda\times e^{-\left[\frac{(x-\mu_x)^2}{2\times\sigma_x^2} + \frac{(y-\mu_y)^2}{2\times\sigma_y^2}\right]} \quad (27)$$

Where λ is the amplitude, μ_x, μ_y are the expected values for *x, y* and σ_x, σ_y are the variances for the variables. The point *(x= μ_x , y= μ_y)* represents the maximum which will be used to find the sub-pixel displacement.

$$f(x,y) = a \times x^2 + b \times y^2 + c \times x \times y + d \times x + f \times y + g \quad (28)$$

Where *a, b, c, d, f* are different parameters of the function and the maximum can be calculated analysing partial derivates.

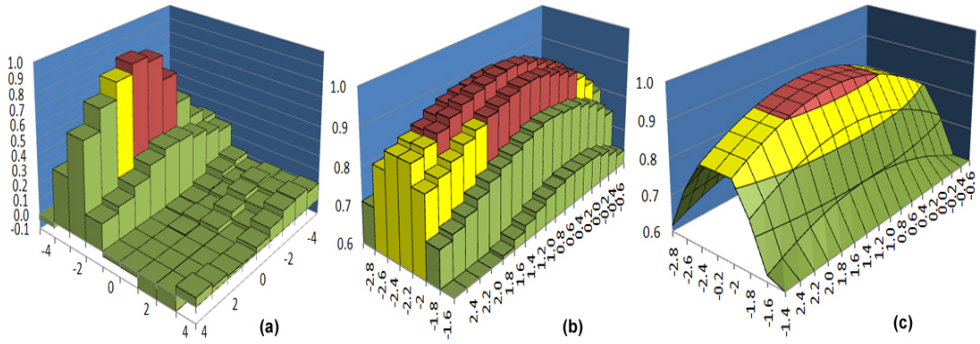

Figure 46: *L-M* subpixel technique.

A similarity peak in a synthetic image sequence is represented. (a) Discrete correlation values, (b) Correlation values near the peak using bilinear interpolation of the image, (c) L-M fitting performed with a Gaussian model using a 3x3 window of discrete values around the maximum.

The Levenberg-Marquardt method can be expressed as follows:

$$\left(J_{n\times m}{}^{T} \times J_{n\times m} + d \times I_m\right) \times Inc_{m\times 1} = J_{n\times m}{}^{T} \times E_{n\times 1}$$

$$J_{n\times m} = \begin{bmatrix} \frac{\partial f(x_1,y_1)}{\partial \mathrm{p}_1} & \frac{\partial f(x_2,y_2)}{\partial \mathrm{p}_1} & \cdots & \frac{\partial f(x_n,y_n)}{\partial \mathrm{p}_1} \\ \frac{\partial f(x_1,y_1)}{\partial \mathrm{p}_2} & \frac{\partial f(x_2,y_2)}{\partial \mathrm{p}_2} & \cdots & \frac{\partial f(x_n,y_n)}{\partial \mathrm{p}_m} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial f(x_1,y_1)}{\partial \mathrm{p}_m} & \frac{\partial f(x_2,y_2)}{\partial \mathrm{p}_m} & \cdots & \frac{\partial f(x_n,y_n)}{\partial \mathrm{p}_m} \end{bmatrix}_{n\times m} \quad (29)$$

Where p_i are the parameters of the function, $d > 0$ is a constant adjusted in each iteration, E_{nx1} is the error matrix calculated using n observations and the corresponding values predicted with the model, J_{nxm} the Jacobian matrix of f, I_m is the identity matrix and Inc_{mx1} is the vector of increments for the next iteration.

An algorithm for updating d is proposed consisting in a modification of the method proposed by Nielsen [145], as shown in (30). According to it, the Residual Sum of Squares (RSS) is calculated from the error matrix (E_{nx1}), so that if the RSS decreases, the d value also decreases, so the algorithm gets close to the Gauss-Newton one; in other instances, the d value increases so that the algorithm behaves as a gradient descent method.

$$
\begin{aligned}
&if\left(\left(RSS^{k-1} < ErrorMin\right)\cup\left(k-1 \geq kMax\right)\cup\left(d > dMax\right)\right) \rightarrow FINISH \\
&else\begin{cases}
Calculate\left(RSS^{K}\right) \\
if\left(RSS^{k} < RSS^{k-1}\right)\ \begin{cases} if\left(d^{k+1} > d^{0}\right)\begin{cases} d^{k+1} = d^{0} \\ v = 2 \end{cases} \\ else\begin{cases} d^{k+1} = d^{k} \,/\, v \\ v = 2 \times v \end{cases} \end{cases} \\
else\begin{cases} d^{k+1} = d^{k} \times v \\ v = 2 \times v \end{cases}
\end{cases}
\end{aligned}
\tag{30}
$$

The initial estimation of the parameters can be also performed in several ways. The *L-M* technique uses three point estimators to obtain a 1D curve and estimate the initial variances, amplitudes or polynomial coefficients. Other parameters such as μ_x , μ_y are simply set to 0.

5.7 Vector Filtering

Once the displacement vectors have been calculated, a vector processing algorithm has been implemented, in which the anomalous vectors are replaced and a soft vector field is obtained.

Thus, the results obtained are more homogeneous, the system's behavior is more stable, and most outliers are ruled out.

The filtering process can be summarized as follows::

1. A vector V, made up with the neighbor blocks of B_i, is defined. Each block B_j in V accomplishes the following criteria: its correlation value c_j is greater than a threshold selected by the user and the distance d_{ij} from B_j to B_i is less than a selected filter radius R.

2. An average displacement $v_{i'}$ is calculated for V, weighting the displacement of each block in V using 2D Gaussian function centered in B_i with a standard deviation defined as follows [75]:

$$\sigma = 0.3 \times \left(\left(R - 1 \right) \times 0.5 - 1 \right) + 0.8 \quad (31)$$

3. For each block B_j in V, the difference $diff_{i'j}$ between its measured displacement and average one is calculated. If $diff_{i'j} > v_{i'}/CF$ and $diff_{i'j} > 2$, B_j is removed from V. Being CF a correction factor selected by the user and set by default to 2.

4. The average displacement for V is updated and v_i is replaced by $v_{i'}$. It should be noted that the current block B_i can be included in V or not according to the explained factors.

5.8 Deformation Models

A deformation model is a transformation which maps all positions in one image plane to positions in a second plane. Therefore, the deformation of an image is primarily a transformation of the plane to itself, where the grey level values are transformed according their associated coordinates.

In this case, the transformation of a set of points (x_i, y_i) which has been calculated using the Block-Matching technique will be used to find a transformation model for each pixel of the image (32).

$$f : R^2 \rightarrow R^2 \quad f(x, y) = (x', y') \qquad (32)$$

Where, the function f must satisfy the conditions $f(x_i, y_i)=(x_i', y_i')$ for the known points.

It should be noted that the terms interpolation and warping are, at this point equivalent concepts. Furthermore, they are also related with filtering and smoothing, since the warping model will be a compromise between a smooth distortion and one which achieves a good match.

In this thesis, different deformation models have been used. They are based on parametric deformation models, Radial Basis Functions (*RBF*) and local parametric models. However, different models such as elastic deformations, or any interpolation technique can be used with the proposed technique.

5.8.1 Parametric Models

Parametric deformation models are easy to compute and need only a few points to be solved. They model simple transformations such as scaling, rotation and linear

displacement. They are useful in many applications where it is important to use a transformation which is no more general than it need be. [34].

In the proposed algorithm, the parametric models can be used as a constraint in the first iterations of the algorithm, to remove outliers and the linear component of the deformation. Then, the search area can be reduced in the next iterations where more general models can also be used. [34].

Figure 47 shows a hierarchy of parametric transformations, ordered from according to their degrees of freedom.

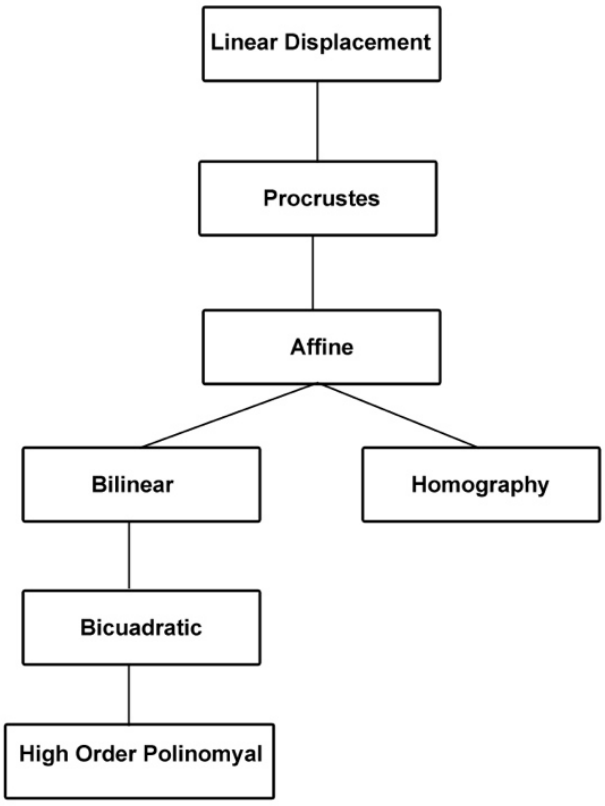

Figure 47: Most common parametric deformation models.

Linear displacement is the traditional model of Block-Matching. Its mathematical formulation can be expressed as follows:

$$\text{Linear Displacement} \quad \begin{cases} x' = x + a \\ y' = y + b \end{cases} \quad (33)$$

The procrustes transform, also called Euclidean transform, uses 4 parameters to model a transformation involving translation, scaling and rotation.

In Computer Vision, location, scale and rotation are known collectively as pose. Once these differences have been removed what remains is known as shape.

The equations of procrustes transform can be expressed as follows.

$$\text{Procrustes} \quad \begin{cases} x' = a + c \times x \times \cos(\alpha) + d \times y \times \sin(\alpha) \\ y' = b - c \times x \times \sin(\alpha) + d \times y \times \cos(\alpha) \end{cases} \quad (34)$$

The affine transformation is a 6 parameter generalisation of the procrustes transformation. It models changes in the shape preserving the collinearity relation between points.

This transformation is modeled with 1st order polynomial equations as follows:

$$\text{Affine} \quad \begin{cases} x' = a + c \times x \times \cos(\alpha) + d \times y \times \sin(\alpha) \\ y' = b - c \times x \times \sin(\alpha) + d \times y \times \cos(\alpha) \end{cases} \quad (35)$$

The homography transformation, or perspective transformation, is a 8-parameter non-linear transformation. It models the projection of a planar object viewed from a point in space. Their equations can be expressed as follows:

$$\text{Homography} \quad \begin{cases} x' = (a_0 + a_1 \times x + a_2 \times y) / (c_0 \times x + c_1 \times y + 1) \\ y' = (b_0 + b_1 \times x + b_2 \times y) / (c_0 \times x + c_1 \times y + 1) \end{cases} \quad (36)$$

The bilinear transformation is another 8-parameter generalization of the affine transformation. It preserves straight lines in three particular directions, including lines parallel to either x- or y-axes. It can be expressed as follows:

$$\text{Bilinear} \quad \begin{cases} x' = a_0 + a_1 \times x + a_2 \times y + a_3 \times x \times y \\ y' = b_0 + b_1 \times x + b_2 \times y + b_3 \times x \times y \end{cases} \tag{37}$$

The bicuadratic transformation is a 12-parameter generalization of the bilinear transformation. It is a 2nd order polynomial transformation expressed as follows:

$$\text{Bicuadratic} \quad \begin{cases} x' = a_0 + a_1 \times x^2 + a_2 \times y^2 + a_3 \times x + a_4 \times y + a_5 \times x \times y \\ y' = b_0 + b_1 \times x^2 + b_2 \times y^2 + b_3 \times x + b_4 \times y + b_5 \times x \times y \end{cases} \tag{38}$$

Finally, a general polynomial transformation can be expressed as follows:

$$\text{Polynomial} \quad \begin{cases} x' = \sum_{i,j} a_{ij} \times x^i \times y^j \\ y' = \sum_{i,j} b_{ij} \times x^i \times y^j \end{cases} \tag{39}$$

5.8.2 Radial Basis Functions (RBF) Models

Radial functions have proven to be an effective tool in multivariate interpolation problems of scattered data [146].

These techniques solve the interpolation-warping problem using a set of radial functions *S* defined in a set of points (x_i, y_i) as follows:

$$\begin{gathered} f : R^2 \to R^2 \quad f(x,y) = (x', y') \\ f(x,y) = \sum_i c_i S\left(\left\| x - x_i, y - y_i \right\|\right) + p(x,y) \end{gathered} \tag{40}$$

Where || || represents the Euclidean distance from the current point to the center of the radial function. Radial basis functions cannot reproduce polynomials, and thus yield a poor approximation with common deformations such as the affine one. To solve this issue a polynomial *p* is added to the formulation.

There are different radial functions with a unique solution for any *n* different points. Some of them have the additional property that the interpolant satisfies some variational principle [147].

In this thesis, the approach proposed by Bookstein [46], was used to estimate the deformation. This technique uses the thin-plate radial function with affine transforms to represent the deformation in a pair of two-dimensional surfaces, and can be expressed as follows:

$$\begin{aligned} f_x(x,y) &= \sum_i c_i S\left(\sqrt{(x-x_i)^2+(y-y_i)^2}\right) + a_0 + a_1 \times x + a_2 \times y = x' \\ f_y(x,y) &= \sum_i c_i S\left(\sqrt{(x-x_i)^2+(y-y_i)^2}\right) + b_0 + b_1 \times x + b_2 \times y = y' \\ S(t) &= t^2 \times \log\left(t^2\right) \end{aligned} \quad (41)$$

The equation expressed above minimises the physical bending energy *J,* which can be expressed as follows:

$$J_f = \int\int_{R^2} \left(\left(\frac{\partial^2 f}{\partial x^2} \right)^2 + 2 \times \left(\frac{\partial^2 f}{\partial x \times \partial y} \right)^2 + \left(\frac{\partial^2 f}{\partial y^2} \right)^2 \right) \times dx \times dy \quad (42)$$

Therefore, to find out the deformation for a given point *(x, y)* using a set of *n* anchor points (x_i, y_i) where the position (x_i', y_i') after deformation is known, the coefficients of the model should be calculated.

The coefficients $c_0, c_1, \ldots, c_N, a_0, a_1, a_2$ to find out x' can be obtained as the solution of a system with $(n+3)$ simultaneous linear equations proposed by Bookstein [46]. This system can be expressed as follows:

$$\begin{aligned} Y_{1\times n+3} &= W_{1\times n+3} \times L_{n+3\times n+3} \\ W_{1\times n+3} &= \begin{bmatrix} c_1 & \ldots & c_n & a_0 & a_1 & a_2 \end{bmatrix}^T \\ Y_{1\times n+3} &= \begin{bmatrix} x_1' & \ldots & x_n' & 0 & 0 & 0 \end{bmatrix}^T \end{aligned} \quad (43)$$

Where L is a matrix defined as follows:

$$L_{n+3\times n+3} = \begin{bmatrix} K_{n\times n} & P_{3\times n} \\ P^T_{n\times 3} & O_{3\times 3} \end{bmatrix}$$

$$K_{n\times n} = \begin{bmatrix} 0 & S(r_{12}) & \ldots & S(r_{1n}) \\ S(r_{21}) & 0 & \ldots & S(r_{2n}) \\ \ldots & \ldots & \ldots & \ldots \\ S(r_{n1}) & S(r_{n2}) & \ldots & 0 \end{bmatrix} \quad (44)$$

$$P_{3\times n} = \begin{bmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ \ldots & \ldots & \ldots \\ 1 & x_n & y_n \end{bmatrix} \quad O_{3\times 3} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Where r_{ij} is the Euclidean distance for two anchor points $||x_i-x_j, y_i-y_j||$.

The same procedure is employed to calculate y'.

Finally, it should be taken into account that the previous formulation can be used to estimate the deformation with different radial functions.

5.8.3 Local Parametric Models

The parametric models can only represent simple types of deformation while more complex models such as *RBF* functions are expensive to compute and are limited

by linear algebra, since in a standard case the solution of hundreds of simultaneous linear equations will be necessary.

The local parametric model proposed, is a mix of parametric models and *RBF* models.

This model takes advantage of the definition of the blocks and uses parametric models in the local context of each block. The general model can be defined as follows:

$$\begin{aligned} f: R^2 \to R^2 \quad f(x,y) &= (x',y') \\ f(x,y) &= \sum_i S_i \times p_i(x,y) \end{aligned} \qquad (45)$$

Where $p_i(x, y)$ is a parametric model depending of m parameters and the S_i is a coefficient taking the following values:

$$S_i \begin{cases} 1/k & \|x - x_i, y - y_i\| = \min\left(\|x - x_j, y - y_j\|\right) \\ 0 & \text{Other case} \end{cases} \qquad (46)$$

Where k is the total number of block centers at the minimum distance to (x, y).

The m parameters of the model for a given block are calculated using the neighbouring blocks to solve a linear equation system.

This procedure has the advantage of allowing the estimation of each block separately, so deformation can be calculated without being limited for complex algebraic models and the process can be easily parallelized.

Figure 48 shows a scheme of how deformation is estimated for a single block using local parametric models.

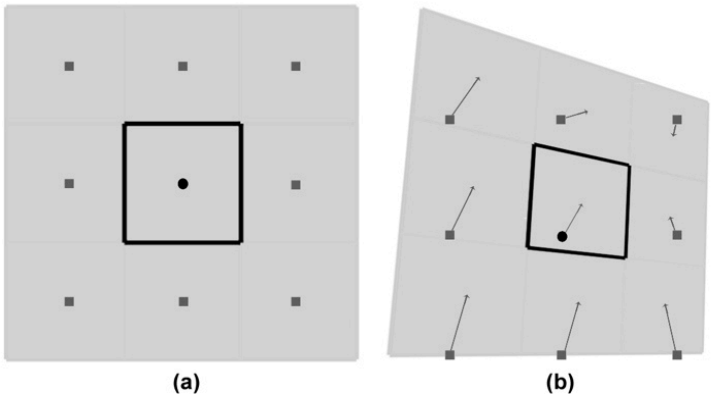

Figure 48: Deformation with local parametric models.

The shape of the block is marked with a black rectangle. Block center is marked with a circle and neighboring block centers with squares. (a) Initial stage: the block has a regular shape. (b) After deformation: displacement has been calculated for each block center. The shape of the block is calculated using a parametric model and the new position of the points.

5.9 Obtaining the Deformed Image

To obtain the deformed image, or a deformed block, using the map of positions calculated with the deformation model, it is necessary to perform an interpolation process on the image. To this end a bilinear interpolation technique, which is the most common model in image interpolation, have been used.

The bilinear interpolation technique is an extension of the linear interpolation for interpolating functions of two variables on a regular grid.

In this case the pixel values in the discrete positions of the image are known and it is necessary to estimate the pixel values in the desired non-discrete positions. The method consists in performing a linear interpolation in the x axis and then another linear interpolation in the y axis.

The procedure to find the intensity value of a pixel *f(x, y),* knowing the intensity in the 4 neighboring pixels $f(x_1, y_1)$, $f(x_1, y_2)$, $f(x_1, y_2)$ and $f(x_2, y_2)$ is shown in Figure 49 and the formulation of the technique is defined in (47).

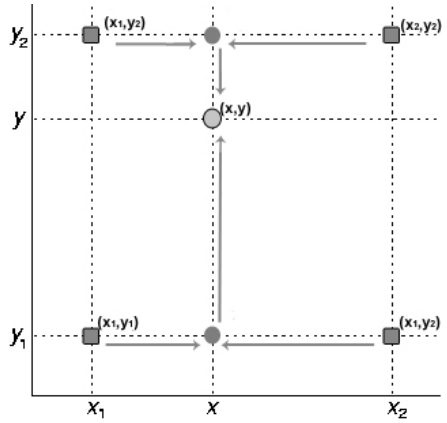

Figure 49: Bilinear interpolation scheme.

$$
\begin{aligned}
f(x,y) &= \frac{y_2 - y}{y_2 - y_1} f(x,y_1) + \frac{y - y_1}{y_2 - y_1} f(x,y_2) \\
f(x,y_1) &= \frac{x_2 - x}{x_2 - x_1} f(x_1,y_1) + \frac{x - x_1}{x_2 - x_1} f(x_2,y_1) \qquad (47) \\
f(x,y_2) &= \frac{x_2 - x}{x_2 - x_1} f(x_1,y_2) + \frac{x - x_1}{x_2 - x_1} f(x_2,y_2)
\end{aligned}
$$

Where *f* is the interpolation function, and $f(x_1, y_1)$, $f(x_1, y_2)$, $f(x_1, y_2)$, $f(x_2, y_2)$ are known values.

5.10 Strain Analysis

After having estimated the displacement, it is necessary to obtain a strain measurement around each point of the object in the test.

We start with a body in an initially non-deformed stage. The effect of strain implies that the body will change its shape, acquiring a different geometry. Thus, a material point, originally in position $P^T=(x, y, z)$, will acquire a new position $P^T=(x',y',z')$ according to displacement u. Here, as we are considering a 2D surface, we only consider a displacement in two dimensions:

$$P' = P + u(x,y) \rightarrow \begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} x + u_x \\ y + u_y \end{bmatrix} \quad (48)$$

Knowing the original position and the deformed one at each point of the body, it is possible to measure the strain effects on a 2D surface:

- Changes to the length between two points in the body
- Changes to the angles between two body fibers.

Thus, 3 standard measurements will be used for strain analysis:

Strain gradient in a point (F_A)

Strain gradient in a point is a matrix that transforms the vectors surrounding that point from the non-deformed configuration to the deformed one:

$$da' = F_A \times da \quad (49)$$

Where da' describes the difference between the new position of the point A and another point B situated at a differential distance of A $(B=A+da)$. After a process of displacement with strain, the points are displaced to new positions A' and B', expressed as follows:

$$\begin{aligned} A' &= A + u \\ B' &= A + u + da' = A' + da' \end{aligned} \quad (50)$$

Being F_A the Jacobian matrix of strain. It is calculated from the matrix G_A known as the displacement gradient:

$$F_A = I_2 + G_A = \left[I_2 + \begin{bmatrix} \frac{\partial u_x}{\partial x_a} & \frac{\partial u_x}{\partial y_a} \\ \frac{\partial u_y}{\partial x_a} & \frac{\partial u_y}{\partial y_a} \end{bmatrix} \right] \quad (51)$$

In order to calculate it, an estimate of the partial derivatives must be performed by using the differences in the measurements of displacements in neighboring blocks.

Specific longitudinal strain (E)

The specific longitudinal strain is relation between the final and the original length of a direction fiber which it is expressed as follows:

$$E = \frac{dl_{A'B'} - dl_{AB}}{dl_{AB}} \quad (52)$$

Where a direction fiber is defined as a segment *AB* where *A* and *B* are two points situated at a differential distance. In practice, two neighboring block centers are used to estimate *E*. Small deformations are usually expressed in microstrain units, $(E \times 10^6)$.

Specific angular strain (H)

This is the change in degrees of the angle between two orthogonal fibers *v* and *u* (α is the initial angle and α' the deformed one). Two orthogonal fibers are defined as a two segments made up of three points *A*, *B* and *C,* situated at a differential distance, where *A* is the common vortex and with an initial angle of 90º. After a time interval, the points are displaced to the positions *A'*, *B'* and *C',* and the fibers

acquire new directions u' and v'. This is expressed in (53). Three neighboring block centers are used to estimate H.

$$\left.\begin{array}{c} \alpha = \dfrac{\pi}{2} = 90^{\circ} \\ \alpha' = \arccos\left(u' \times v'\right) \end{array}\right\} \rightarrow H = \alpha' - \alpha \qquad (53)$$

6 Experimental Results

6.1 Comparative Tests

There are several datasets and methodologies used in the literature to evaluate the performance and accuracy of Optical Flow algorithms, being the most important the Middlebury evaluation [148].

The main problem to validate the current algorithm is that, most current datasets don't deal with deformations and the used image sequences usually represent a scene with multiple rigid motions. In particular, the Middlebury ranking is focused in this kind of motion and there are no Block-Matching algorithms in it.

In order to evaluate the accuracy and performance of the presented technique, different sets of experiments were performed using several image sequences.

The composition of the dataset is as follows:

- The *Sphere* and *Cube and Sphere* (with complex object motion) sequences (Figure 50). These are some of the sequences granted by Otago University, New Zealand [140] and used in different *CV* works [149]. These are examples

of sequences of objects with simple boundaries with a 3D rigid motion. These sequences have a size of 200x200 pixels.

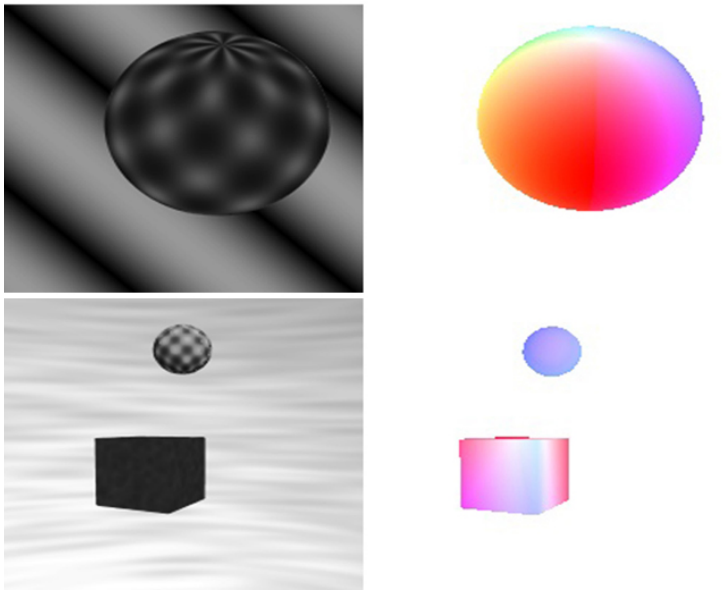

Figure 50: The *Sphere* and *Cube and Sphere* sequences.

With each sequence the ground truth is represented, where the intensity value represents the magnitude and the color represents the direction of the displacement in each.

- The *Yosemite* and *Yosemite with clouds* landscape sequences (Figure 51). The *Yosemite* sequence is a standard test for benchmarking optical flow algorithms. It was created by Lynn Quam [150]. It contains a diverging field, occlusions and multiple motions at the horizon and it was widely studied in different works [151]. The *Yosemite with clouds* is a variant of this sequence, with a fractal cloud pattern with Brownian motion in the horizon. In the cloud region of this sequence the ground truth is an estimation and the assumption of brightness constancy does not hold. The *Yosemite* landscape sequences represent camera motion, but, as there are no severe discontinuities in the motion field, the optical flow of the terrain can be described as a deformation

process, being this sequence ideal for a Block-Matching analysis. These sequences have a size of 316x252 pixels.

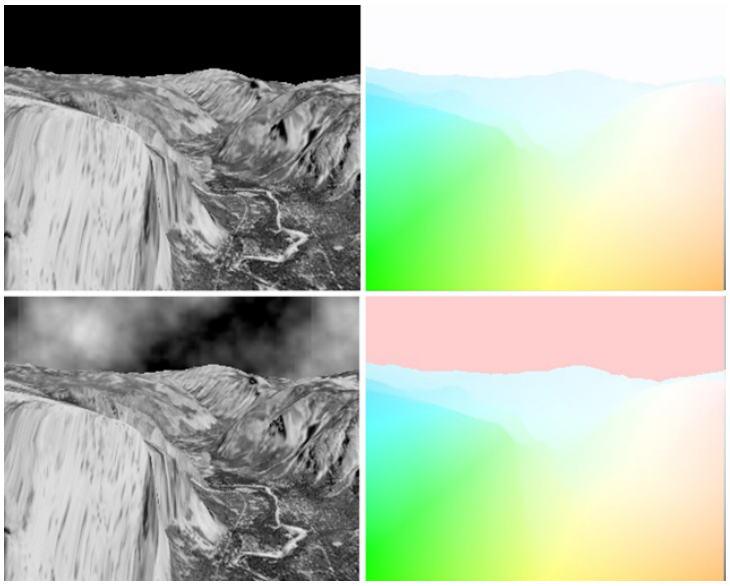

Figure 51: The *Yosemite* and *Yosemite with clouds* sequences.

With each sequence the true ground truth is represented, where the intensity value represents the magnitude and the color represents the direction of the displacement in each point.

- The *Groove2* and *Urban2* synthetic sequences (Figure 52) granted by the Middlebury University [148] used to train and evaluate optical flow algorithms in several works [152-154]. These sequences represent camera motion in a 3D scene with high complexity boundaries and they are very difficult to analyse with Block-Matching techniques. Theses sequences have a size of 640x480 pixels.

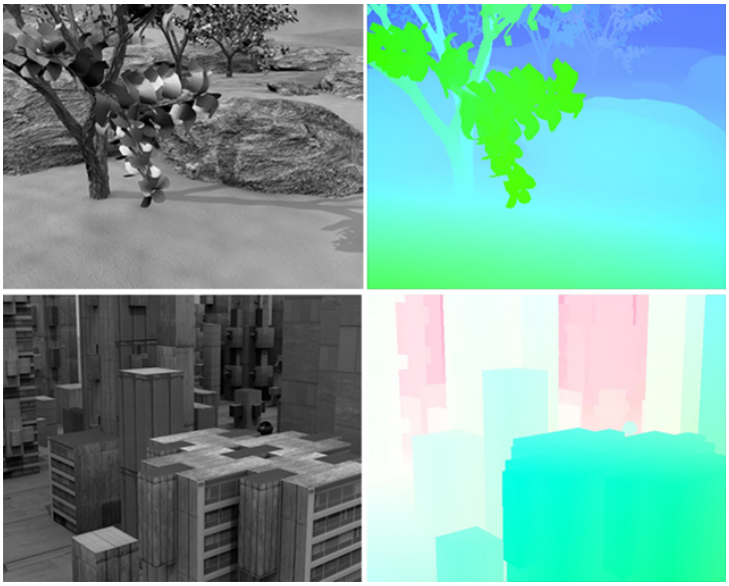

Figure 52: The *Groove2* and *Urban2* sequences.

With each sequence the true ground truth is represented, where the intensity value represents the magnitude and the color represents the direction of the displacement in each point.

- The *Dimetrodon*, *Hydrangea* and *Rubber Whale* (Figure 53) real sequences, they are part of the Middlebury training set, [148]. And represent different kinds of rigid motions. These sequences have a size of 584x388 pixels.

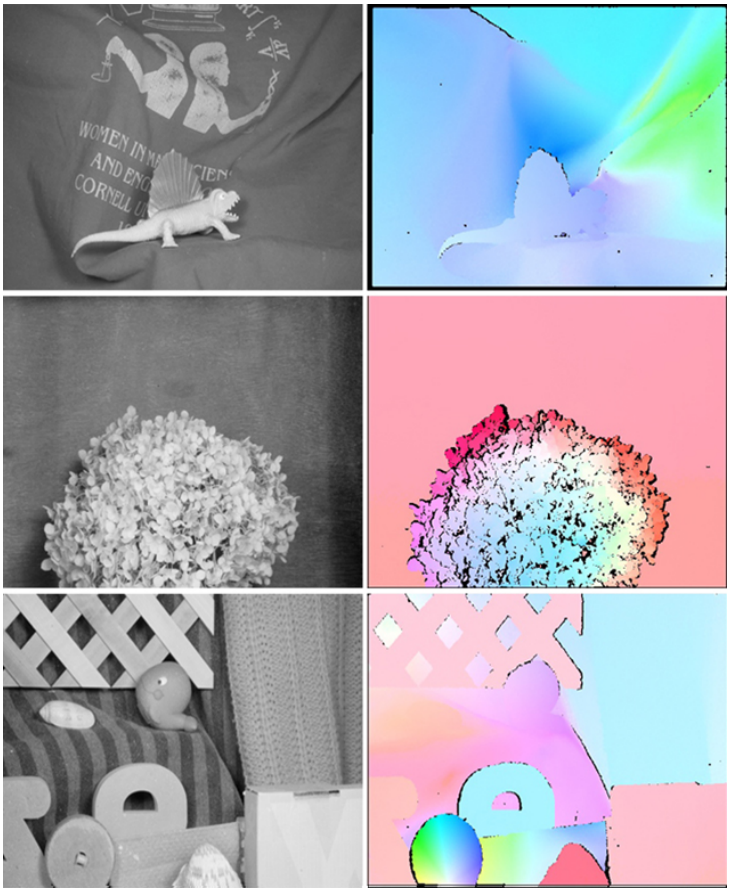

Figure 53: Sequences of real images used in the test.

With each sequence the true ground truth is represented, where the intensity value represents the magnitude and the color represents the direction of the displacement in each point.

This tests were performed analysing the motion between two consecutive frames, selecting full image region, calculating the error statistics according to the true ground data, and using the metrics and methodologies published by Baker et al. [152]. The endpoint (*EP)* error metric defined in (54) was used to measure the accuracy of the techniques.

$$EP = \sqrt{\left(u_M - u_T\right)^2 + \left(v_M - v_T\right)^2} \qquad (54)$$

Where (u_M, v_M) is the measured flow in a point and (u_T, v_T) is the true flow. The *EP* error metric represents the Euclidian distance with the real displacement and it is more relevant than the Angular Error (AE) in practical fields, such as engineering or medicine. Furthermore, the *AE* is not a desirable metric in the aim of this application because it penalizes errors in large flows less than in small flows without a justified reason [152].

6.1.1 Comparative with different configurations of the proposed technique

In order to evaluate the accuracy and performance of the algorithm, the image sequences were analysed with different configurations of the technique. The purpose of this assay is to compare the results obtained with different similarity metrics, iterations, sub-pixel methods or deformation models.

Therefore, a set of experiments were conducted to evaluate the influence of different parameters of interest. To this end, a base configuration was empirically selected and one or more parameters were varied in each assay using OFAT (one factor at a time) and factorial experimental designs.

Table 3: Base configuration for the comparative tests

Configuration Parameters	
Blocks	Blocks of 15x15. Distributed in a regular grid with a separation of 11x11
Similarity metric	Pearson Correlation
Iterations	7 (Without pyramidal decomposition)
SubPixel Method	L-M with 2D Gaussian model and 500 max iterations.
Filtering	Filter Radius: 60 (1st It.), 45 (Middle It.), 35 (Last It.) Minimum Correlation: 0.5 Correction Factor : 2
Deformation Model	LP. Homography + RBF Thin Plate (Last It.)
Dense Filtering	5x5 2D Gaussian filtering

The base configuration used for every assay is detailed in Table 3.

In the first experiment, the impact of block size and block density in the accuracy of the results is analysed.

Since the impact of any feature analysis technique to select block centers shall be strongly dependent of the type of images, and therefore difficult to analyse, in these experiments a block distribution in a regular grid is selected.

The experiment is conducted using a factorial design and selecting the block radius and the separation between blocks as independent variables and measuring the variation in the *EP* error.

Analyzing the obtained results (Figure 54) the election of the block size and density has a mayor impact in the accuracy and, although they are not independent from each other. Independence can be assumed for experimental design in non-extreme values.

Generally, more accurate results can be obtained using bigger blocks and a greater block density. However increasing the size and density of the blocks will increase the computation time and increasing the block size beyond a certain point will increase again the error.

The optimum value for these parameters is related with the complexity of the movement and the resolution of the image. However, the results are not very sensitive to small variations and a wide range of configurations can normally be used.

A good configuration may be from 11x11 to 15x15 blocks for small or middle sized images, although when analysing high resolution images or soft deformations, much bigger blocks can be used.

The distance between blocks is related with the grain of detail that can be measured, the complexity of the motions and the size of the different shapes. Being necessary few blocks to measure rigid motions but being required a high density of blocks for movements with strong deformations.

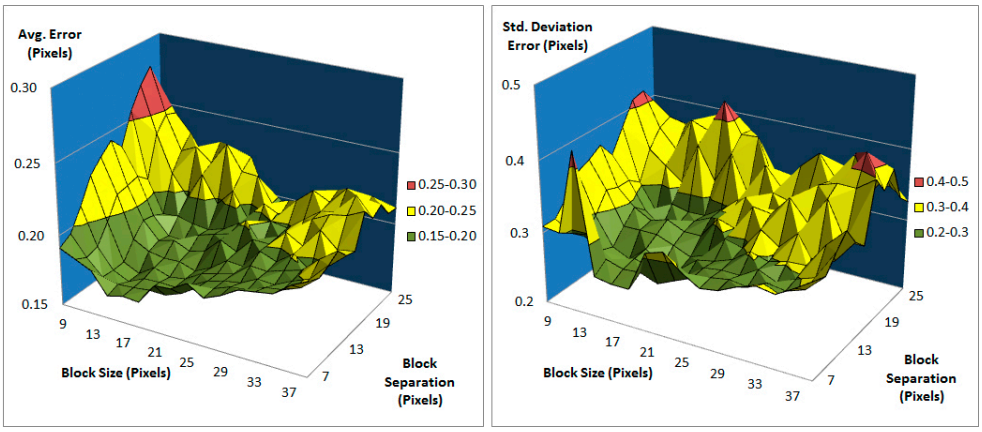

Figure 54: Average accuracy according to block size and block density.

In the second experiment, the impact of filtering in the accuracy of the results is analysed. To this end an OAF experimental design is used using the filter radius (normalized with the block separation) as independent variable.

The obtained results can be seen in Figure 55.

As it can be observed, the use of filtering improves significantly the accuracy of the results while increasing the filter radius causes small effects.

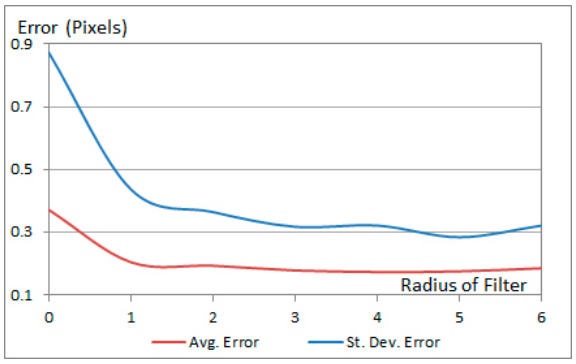

Figure 55: Average accuracy according to the filtering step.

In the third assay, the different proposed metrics were compared. The used experimental design is an OFAT strategy selecting the variable similarity metric as independent value.

Analysing the obtained results (Figure 57) it can be seen that *MSE*, Pearson and *CSM* achieved good results. While the *CSM* obtained more robust results in some sequences, I has not achieved a significant advantage over the Pearson metric and additionally, the adjustment of the k parameter is not easy to justify.

MI similarity, having higher computational requirements, obtained less precise measurements and should be used only in multimodal images.

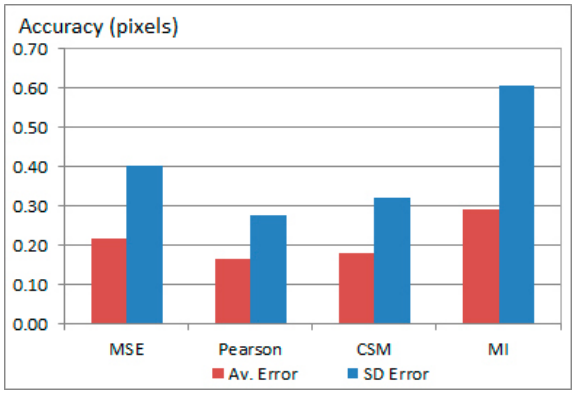

Figure 56: Average according to the similarity metric.

In the fourth assay the different proposed sub-pixel methods were compared, using the same strategy than in the previous assay.

Analysing the results from Figure 57, it can be seen that the use of sub-pixel techniques improves the result in a very significant way.

In the performed assays, the Gaussian models, obtains more accurate results and the L-M technique improved the 3-Point estimators, although it requires more computational time.

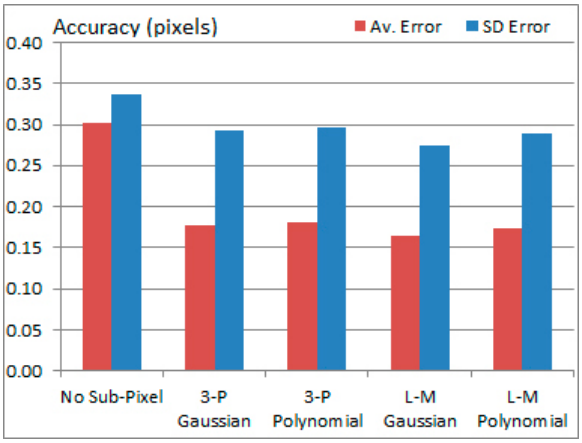

Figure 57: Accuracy and sub-pixel techniques.

In the last assay, the parameters related with deformation are analysed. To this end, the proposed technique was tested with different iterations, different deformation models and activating and deactivating the multilayer processing.

Four different configurations of deformation were used: A local parametric model using homographies, a local parametric model using affine transformations, a Thin Plate RBF model and a mixed approach using a Thin Plate model in the last iteration and a local parametric model in the other ones.

The experiment was conducted with a classic factorial design performing 120 executions with the whole dataset.

Analysing the obtained results in Figure 58, it can be seen that that the results are significantly improved by using an iterative approach. With this technique, the average error and the deviation of the error are reduced and then stabilized after a few iterations. It also can be seen that the use complex models of deformation such as the Thin Plate RBF, does not get a significant improvement over the much simpler local parametric one.

Finally, it can be seen that the use of a segmentation step with a multilayer analysis achieves a significant and constant improvement regardless the deformation model used.

Therefore, it can also be concluded that the three analysed parameters are independent from each other, being the convergence of the solution independent of the deformation model and both parameters independent of the use or not of a segmentation step.

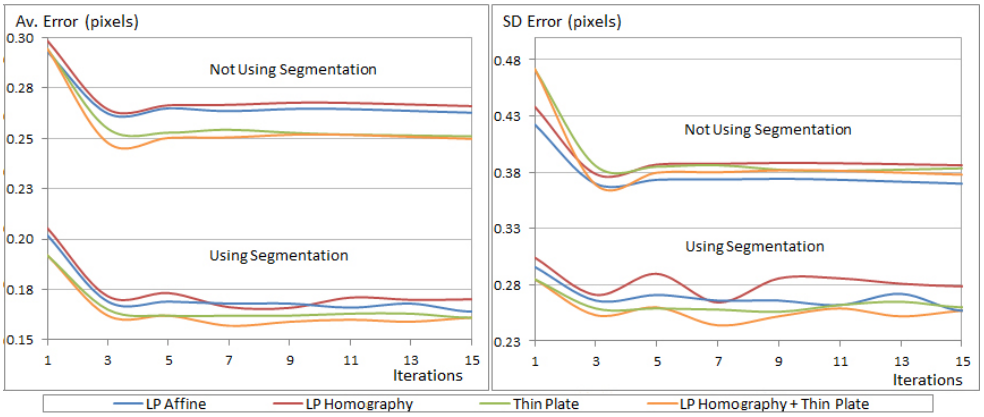

Figure 58: Accuracy and deformation parameters.

6.1.2 Comparative with other algorithms

The purpose of these experiments is to compare the results obtained with the proposed algorithm with different state of the art techniques:

- The *Block-Matching* technique provided by the computer vision library *OpenCv* available at [75].
- The *DaVis* system, from *LA Vision* [155]. A commercial application using Block-Matching for measuring strain in structures. Its algorithm was introduced by [98], and enhanced by [99]. It has been widely used in publications and experiments in various fields [156].
- Modern implementations of the classic *Black & Anandan* and *Horn & Schunck* techniques carried out by [153]. These implementations can be found in [157].
- The *Classic+NL* technique carried out by [154] and available at [157].
- A *Variational* optical flow technique carried out by [158], extended in [159] and available at [160].

In this experiment, the configuration used for the Block-Matching techniques was as close as possible to the one used with the proposed technique. The configuration for the rest of the techniques was the one suggested by their respective authors, using the default parameters when possible which are supposed optimal to the Middlebury training set [148].

Analyzing the results of the experiment (Figure 59), it can be seen that the proposed technique and the *Classic+NL* one achieved the lowest average error of the table. Additionally the proposed technique obtained the lowest standard deviation of the error, obtaining the best results of the tested techniques.

In the other hand, the *DaVis* technique obtained better results than the *Block-Matching* technique from *OpenCv*, but both techniques obtained worse results that the other algorithms.

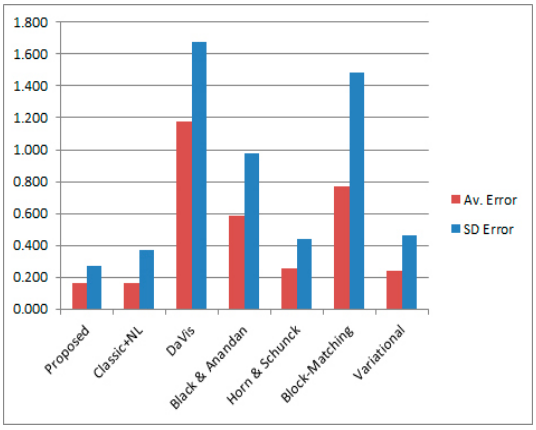

Figure 59: Accuracy with the different *CV* techniques.

Table 4: Average error from the comparative test with synthetic data.

The results showed for the *DaVis* technique were calculated using only the block centers (since this technique does not use a dense displacement field).

	Av. Error (pixels)						
	Proposed	Classic +NL [154]	DaVis [99]	Black & Anandan [153]	Horn & Schunck [153]	Block-Matching [75]	Variational [159]
Yosemite	0.081	0.134	0.201	0.187	0.291	0.483	0.218
Yos. Clouds	0.046	0.353	0.277	0.314	0.406	0.957	0.312
Sphere	0.094	0.097	1.611	0.098	0.324	0.371	0.238
Cube & Sphere	0.326	0.046	0.259	0.072	0.115	0.274	0.225
Groove2	0.174	0.150	0.324	0.174	0.196	0.668	0.168
Urban2	0.176	0.271	7.201	3.984	0.471	2.004	0.453
Dimetrodon	0.224	0.134	0.173	0.145	0.216	0.987	0.173
Hydrangea	0.136	0.192	0.305	0.191	0.192	0.563	0.188
Rubber Whale	0.228	0.104	0.263	0.114	0.118	0.634	0.152
AVG	0.165	0.165	1.179	0.587	0.259	0.771	0.236

Table 5: St. Dev. error from the comparative test with synthetic data.

The results showed for the *DaVis* technique were calculated using only the block centers (since this technique does not use a dense displacement field).

	SD Error (pixels)						
	Proposed	Classic +NL [154]	DaVis [99]	Black & Anandan [153]	Horn & Schunck [153]	Block-Matching [75]	Variational [159]
Yosemite	0.171	0.132	0.278	0.201	0.261	1.070	0.248
Yos. Clouds	0.168	0.445	0.297	0.388	0.407	1.644	0.403
Sphere	0.098	0.207	2.481	0.202	0.418	0.621	0.350
Cube & Sphere	0.399	0.146	0.283	0.218	0.251	1.239	0.389
Groove2	0.311	0.490	0.556	0.460	0.405	1.111	0.440
Urban2	0.155	0.967	10.031	6.447	1.359	3.960	1.416
Dimetrodon	0.425	0.143	0.182	0.146	0.214	1.173	0.166
Hydrangea	0.227	0.448	0.508	0.418	0.364	1.350	0.423
Rubber Whale	0.512	0.357	0.442	0.310	0.266	1.168	0.364
AVG	0.274	0.371	1.673	0.977	0.438	1.482	0.466

The accuracy of the proposed technique can be visually analysed in Figure 60, Figure 61, Figure 62 and Figure 63 where measured displacement for each sequence is shown with the corresponding ground truth.

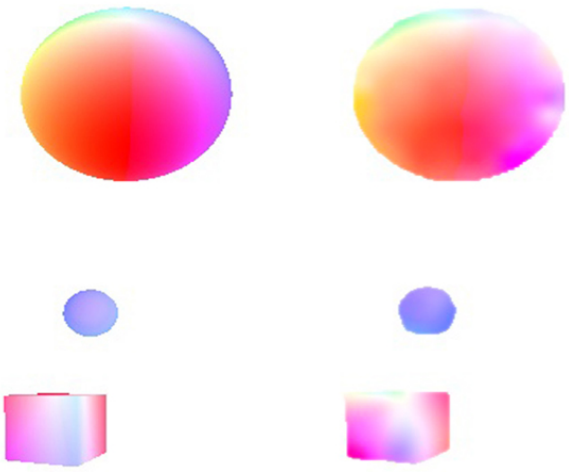

Figure 60: Results in the *Sphere* and *Cube and Sphere* sequences.

The ground truth is represented in the left and the estimated displacement in the right. The intensity value represents the magnitude and the color represents the direction of the displacement.

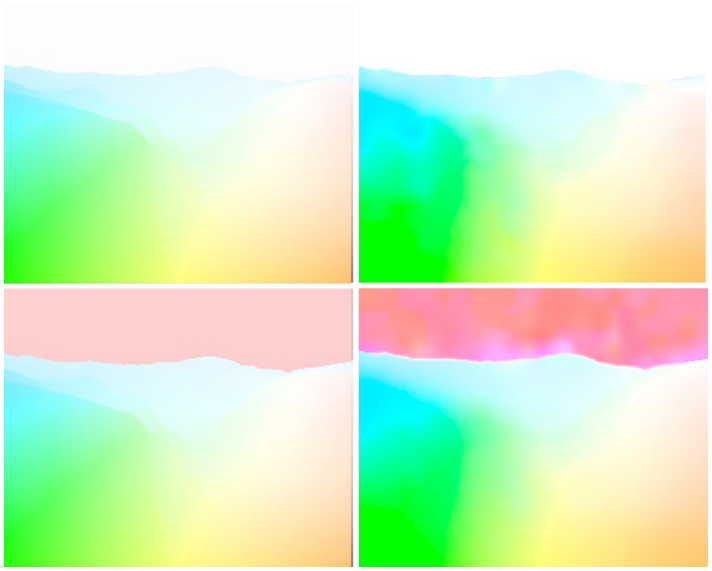

Figure 61: Results in the *Yosemite* and *Yosemite with clouds* sequences.

The ground truth is represented in the left and the estimated displacement in the right. The intensity value represents the magnitude and the color represents the direction of the displacement.

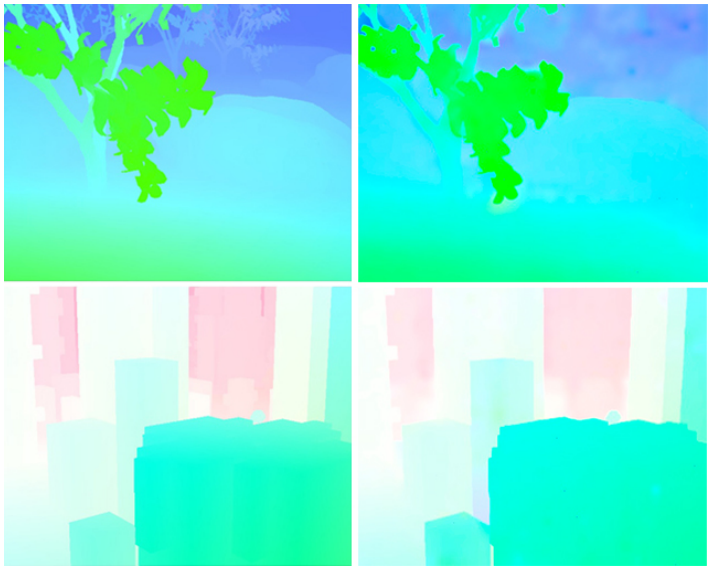

Figure 62: Results in the *Groove2* and *Urban2* sequences.

The ground truth is represented in the left and the estimated displacement in the right. The intensity value represents the magnitude and the color represents the direction of the displacement.

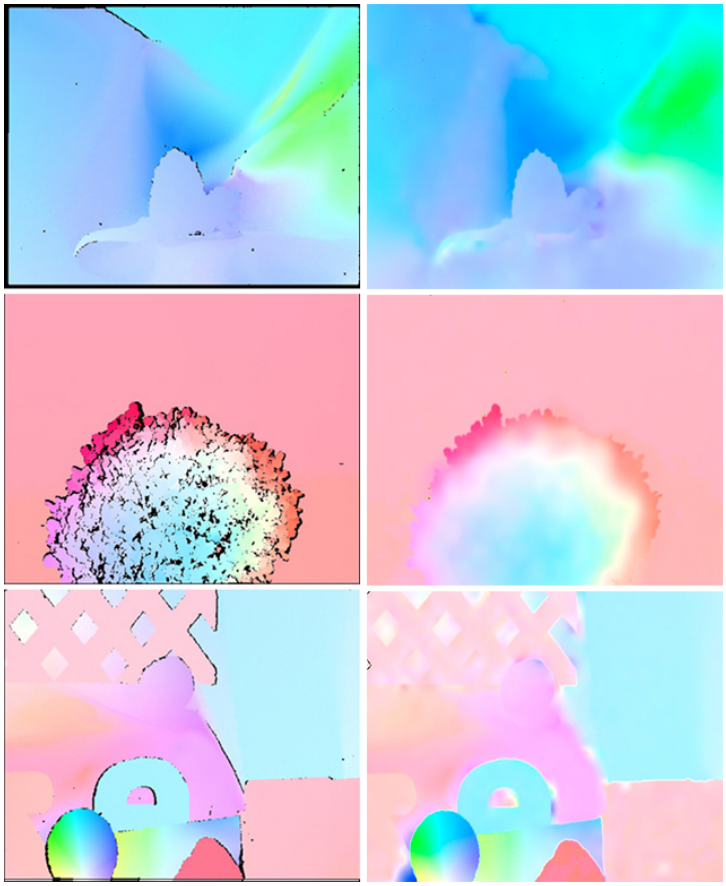

Figure 63: Results in the sequences of real images used in the test.

The ground truth is represented in the left and the estimated displacement in the right. The intensity value represents the magnitude and the color represents the direction of the displacement.

6.2 Analysis of real images without strain

In order to measure the system accuracy in a real scenario, an experimental assembly was carried out, setting a simulated material to a stand at a given distance from a conventional digital camera which took pictures of the object. The camera focus, zoom and color parameters were set up manually.

In this test the used lightning used was a combination of natural (variable and unstable) and artificial light. This reproduces the common luminosity conditions during a strength test. The simulated material used has the visual texture of a natural granitic stone.

In this test the *FOV* was 80mm and the resolution of the used images was 10 megapixels *(mm/pixel = 0.02)*.

The execution of this experiment is illustrated in Figure 64.

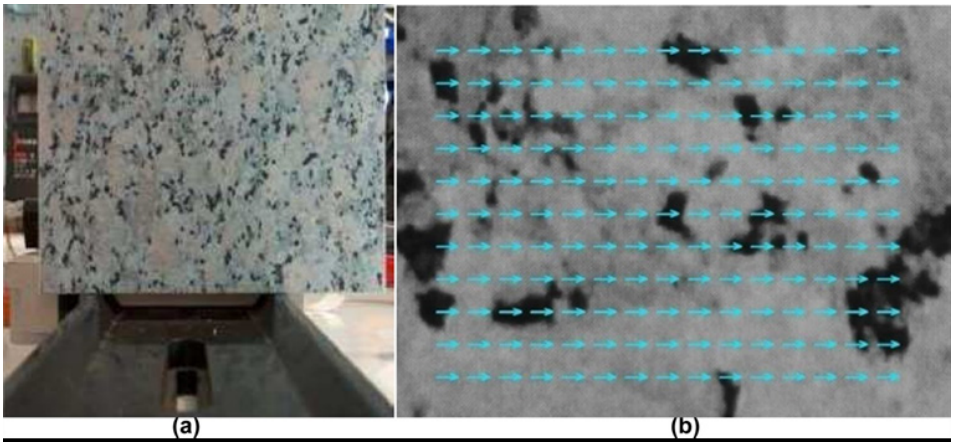

Figure 64: Experiment with simulated material.

Images from an experiment with real images and simulated material, showing a picture taken of the material (a) and the results obtained by the algorithm (b).

This experiment was analysed with the presented algorithm and the results were compared with those obtained with the available techniques from the previous assays (except the *Variational* technique which was not available when these assays were performed).

The results were calculated comparing every frame of the sequence with the initial one (since, as explained before, most of the techniques cannot manage measurements in image sequences). The measurements have been performed in

millimeters, using a calibration pattern in order to correct the distortion caused by the lens and to obtain a measurement on a real scale.

A first experiment was performed with the object in a static position, and assuming that the value of displacement in each object point with regard to the camera was null. Thus, the measurements provided by the system will allow the calculation of the maximum accuracy of the system in an optimistic scenario in real conditions.

Using the same initial conditions, a second test was performed, in which the material was displaced 0.1 millimeters per frame using a micrometer (a device incorporating a micrometric screw for causing high precision displacements to a material). Those displacements were contrasted with a dial depth gauge (a Civil Engineering tool for measuring displacement of a material to which said device has been attached) and used as ground truth to estimate the error of the system.

The execution of these experiments is illustrated in Figure 65.

Figure 65: Execution of the experiment with real images and no strain.

A simulated material was placed in a stand and then displaced using a micrometer and a dial depth gauge was employed to contrast the measurements. A standard digital camera was used to record the test. The mean of the analogical measurements was used as ground truth.

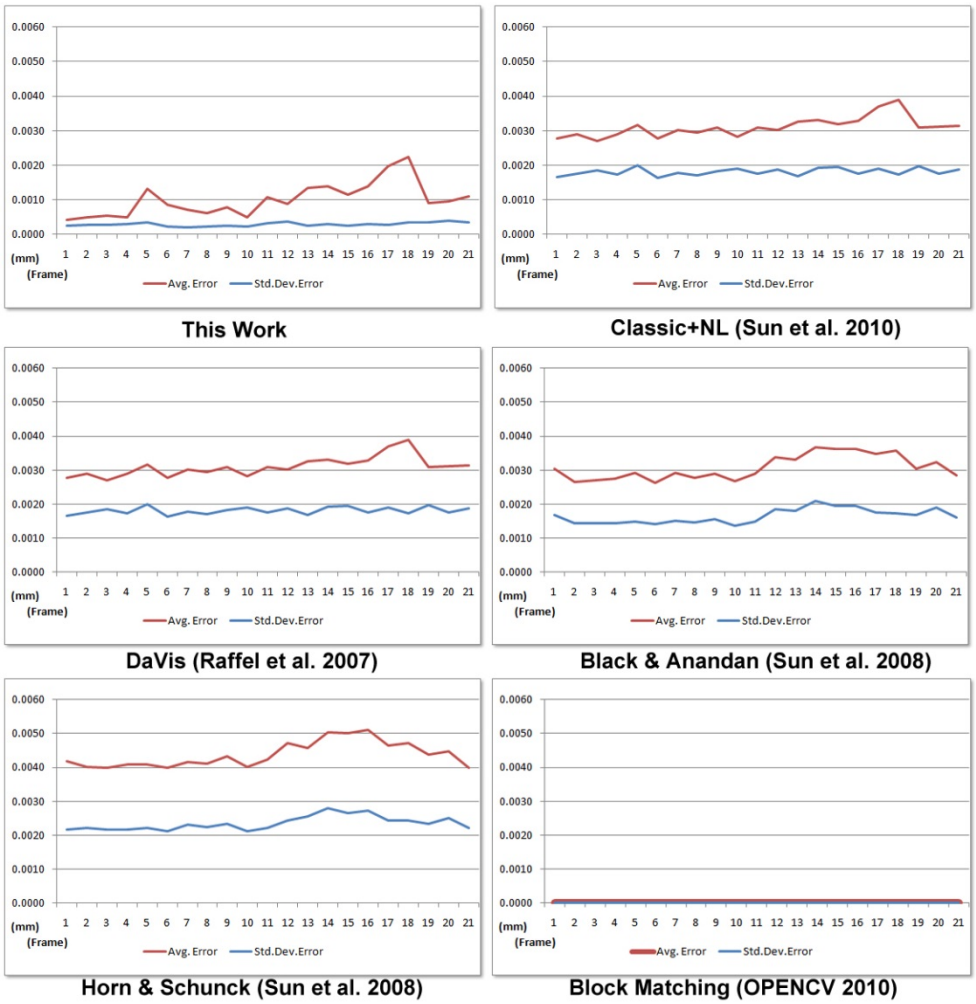

Figure 66: Results of the static experiment with simulated material.

Evolution of the Average Error and Average Standard Deviation of Error expressed in millimeters with the different tested techniques in the static test. Note: The Block-Matching algorithm did not perform any measurement.

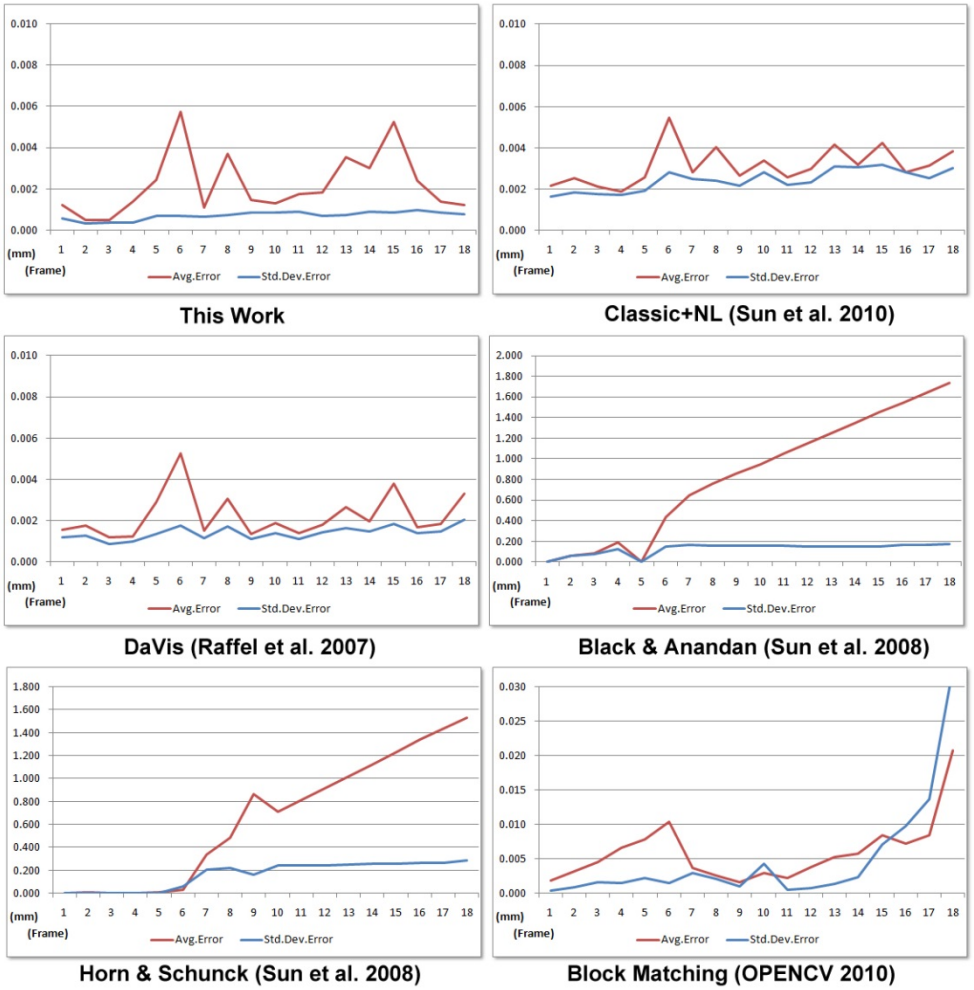

Figure 67: Results of the displacement experiment with simulated material.

Evolution of the Average Error and Average Standard Deviation of Error expressed in millimeters with the different tested techniques in the test with displacement.

The charts in Figure 66 and Figure 67 show the evolution of the average and standard deviation errors in the experiments.

The global results may be clearly observed in Table 6.

Table 6: Results of experiments with real images of a simulated material.

[1] Not enough precision to measure any displacement.

	Static Experiment		Displacement Experiment	
	Global Avg. Error (mm)	Global Std.Dev.Error (mm)	Global Avg. Error (mm)	Global Std.Dev.Error (mm)
This application	**0.0010**	**0.0003**	**0.0022**	**0.0007**
DaVis [99]	0.0018	0.0009	0.0022	0.0014
Classic+NL [154]	0.0031	0.0018	0.0032	0.0024
Block-Matching [75]	0.0000[1]	0.0000[1]	0.0059	0.0048
Black & Anandan [153]	0.0031	0.0017	0.8421	0.130
Horn & Schunck [153]	0.0048	0.0024	0.6579	0.1644

Analysing the obtained results, it may be noted that the ranking of the results with the Static experiment was the same as the one obtained with the synthetic sequences, in which all the algorithms, have obtained good results.

It should be noted that the *Block-Matching* technique does not possesses a sub-pixel precision and did not detected any change in the static test.

In the Displacement Experiment as it can be seen in Figure 67, the *Black & Anandan* and the *Horn & Schunck* algorithms failed when displacements were long enough and the *Block-Matching* algorithm managed to measure the global displacement field while obtaining a very bad precision.

To analyse the obtained results it must also be taken into account that it is not possible to differentiate between system errors and those produced by analogical measurements in the experiment with displacement. In particular, the dial depth gauge and the micrometer allow measuring and applying displacements with a precision not higher than 3 microns.

Therefore, the peaks of the variation in the average error (very similar in every good-performing algorithm) may be attributed to the analogical instruments. Therefore, it may also be concluded that the algorithm presented, together with the *DaVis* and *Classic+NL* ones, was significantly more accurate and stable than the analogical tools used as reference.

The best results in this experiment have been obtained by the *DaVis* system and the proposed technique.

It may be concluded that the proposed technique can obtain very accurate results with real images. Additionally, the measurements obtained in the performed tests were similar in accuracy and better in stability than the analogical tools used as reference.

Furthermore, the accuracy in the proposed system is linked to the pixel level, so, theoretically it is limited only by the zoom and the resolution of the camera used.

6.3 Analysis of real strength tests

Different tests in a scenario with real strains have been performed to analyse the algorithm behavior in the considered application field.

The goal will be to determine the validity and potential application of the algorithm in a real scenario. For this purpose, tree strength tests have been carried out with different materials and forces.

- In the first tests, a steel bar under tensile forces was studied. In this test the obtained results were compared with those obtained by the *CV* techniques used in the previous tests.

- In the second test a concrete test model under compression forces was analyzed. In this test the proposed technique was compared with data obtained with standard strain gauges.
- In the last experiment, an aluminum bar under tensile forces was tested. Obtained results were compared with those provided by a contact extensometer.

These tests were carried out at the Center of Technological Innovation in Construction and Civil Engineering (*CITEEC*) and they are standard tests in the Civil Engineering field.

The materials painted with a random spot distribution to provide a visual texture.

6.3.1 Real strength tests with a steel bar

In the first experiment, the behaviour of a *200 mm (*length*) x 32mm* (diameter*)* steel bar used as reinforcement of structural concrete will be analyzed. The material was subjected to tensile forces up to failure.

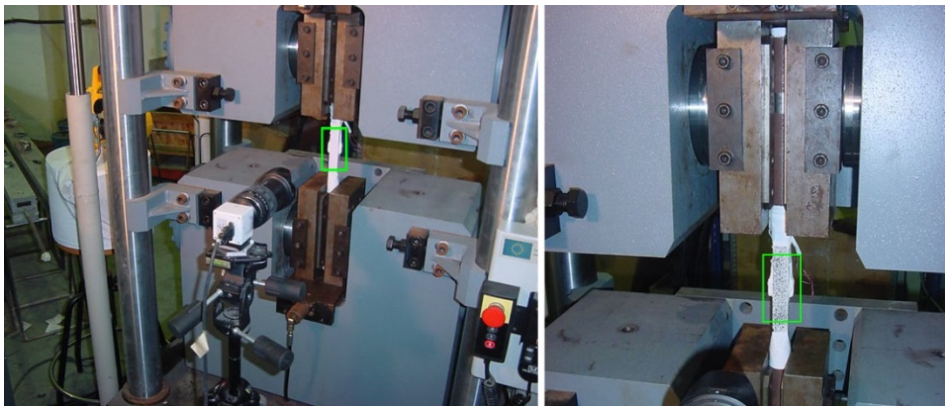

Figure 68: Execution of the tensile test.

Pictures of the tensile strength tests with a steel bar performed at the Center of Technological Innovation in Construction and Civil Engineering (CITEEC). The marked rectangle represents the steel bar (recorded area).

The execution of this experiment is shown in Figure 68.

Figure 69 and Figure 70 illustrate how the material strain develops and the obtained output by the algorithm proposed. Figure 69 shows the displacement field in the complete processed area and Figure 70 shows a 2D representation of the deformation in the shape of the material during the test.

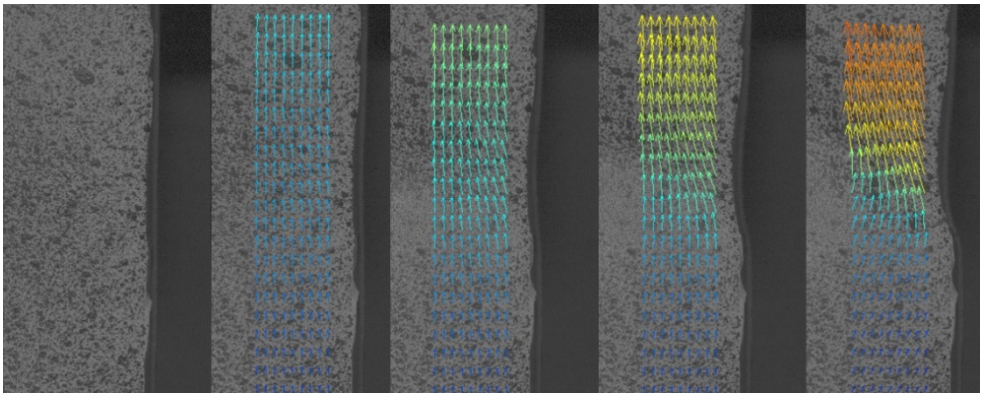

Figure 69: Displacement obtained in the tensile test.

Different images of the sequence are shown with calculated displacement vectors.

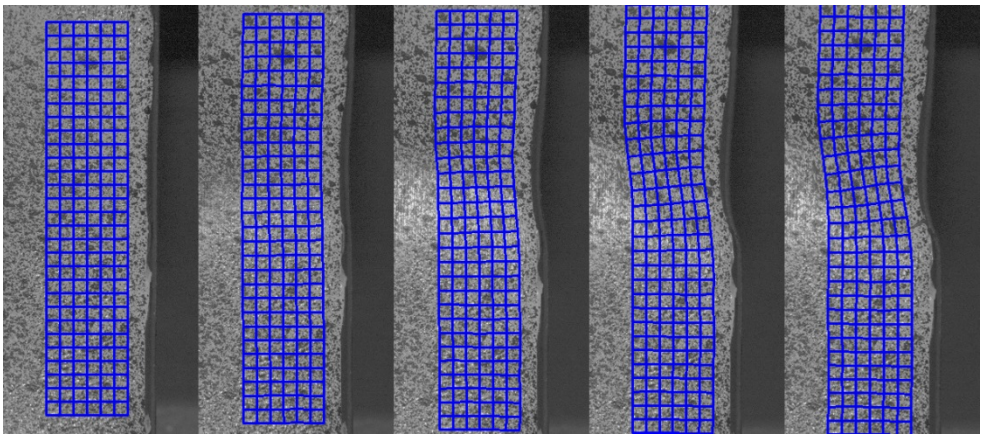

Figure 70: Strain in the tensile strength.

Different images of the sequence are shown with a virtual deformable grid linked to the surface of the material.

The obtained results were compared with those obtained by the *CV* techniques from the previous tests in a qualitative and a quantitative way.

Although, as explained before, only the *DaVis* technique and the proposed one are suitable to perform measurements in long image sequences, it is still possible to compare the goodness of each algorithm when measuring deformations by comparing each image of the sequence with the initial one.

In this test the *FOV* was 90mm and two different analyses were performed.

In the first one the original resolution of the sequence (4 megapixels) was used, obtaining an *mm/pixel* relation of *0.04*.

In the second analysis the sequence was downsampled to 1 megapixel *(mm/pixel = 0.11)* and some compression artifacts were added to increase the effects of the deformation and to simulate a recording with less image quality. Additionally, it was included a stage of vertical displacement prior to the application of the load to obtain long range displacements (which may be more difficult to detect).

To estimate the ground truth data, the images were oversampled to 4 times their original size and a set of points were selected. Then, the displacement of those points was manually estimated. Finally, calibration data was used to transform the obtained measurements into a real scale. This procedure is showed in Figure 71 for a sample point.

Considering a precision of 1 pixel in the manual estimation of the displacement, it can be assumed that the maximum accuracy of the ground truth is close 5 μm (similar to a high precision dial depth gauge).

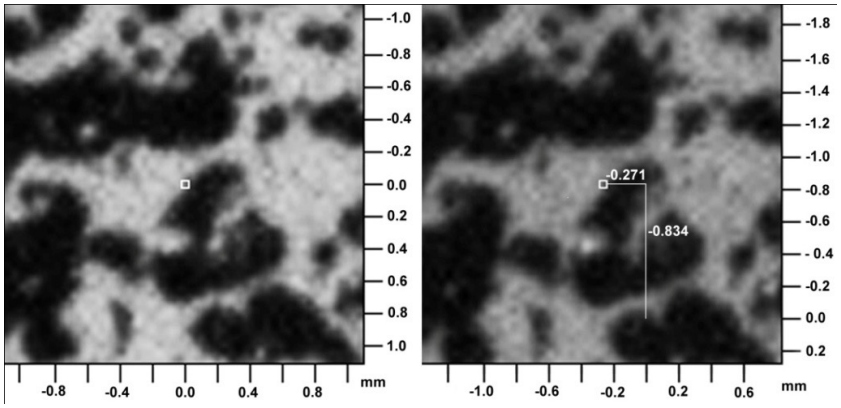

Figure 71: Estimation of displacement for a single point.

First, the selected point was marked with a white rectangle in the first frame. Then, the correspondence for the selected point is calculated in other frame of the sequence. Scale and displacement are shown in millimeters using the calibration data.

The results of this experiment can graphically observed in Figure 72 and Figure 73, and numerical results are shown in Table 7.

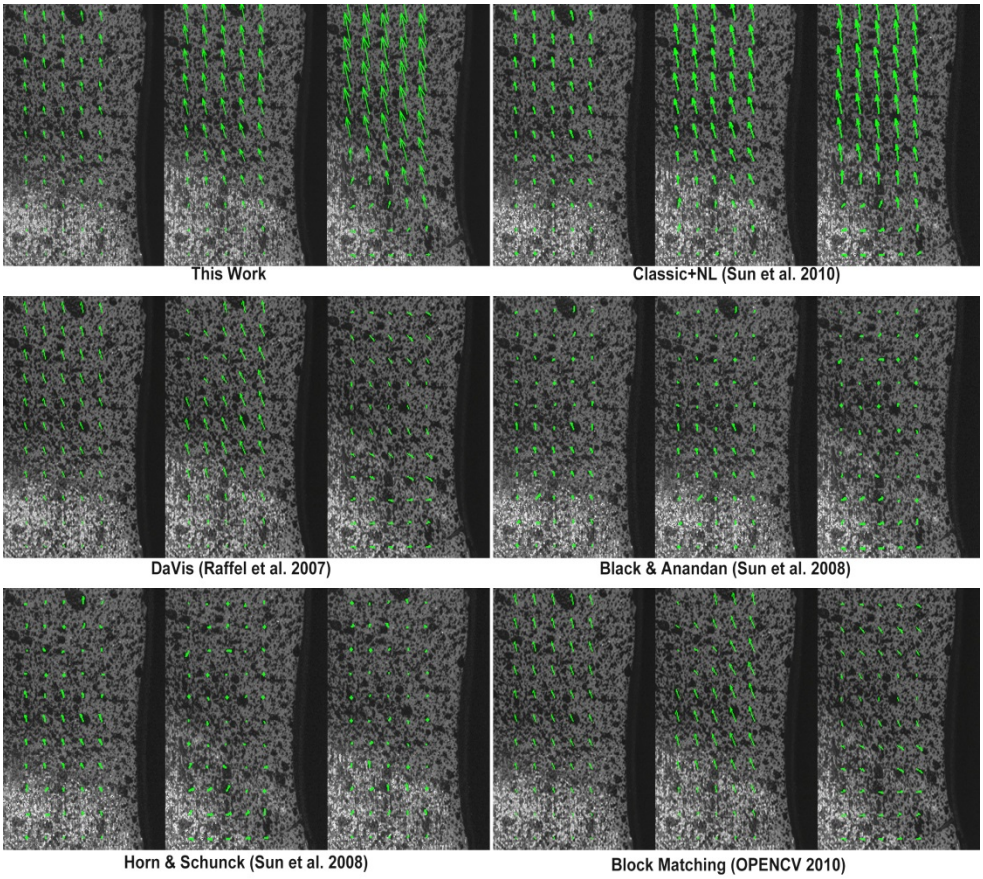

Figure 72: Visual results with the different techniques in the tensile test.

The same three images are shown for each technique. They represent the enlargement of an area of the material with the corresponding displacement vectors.

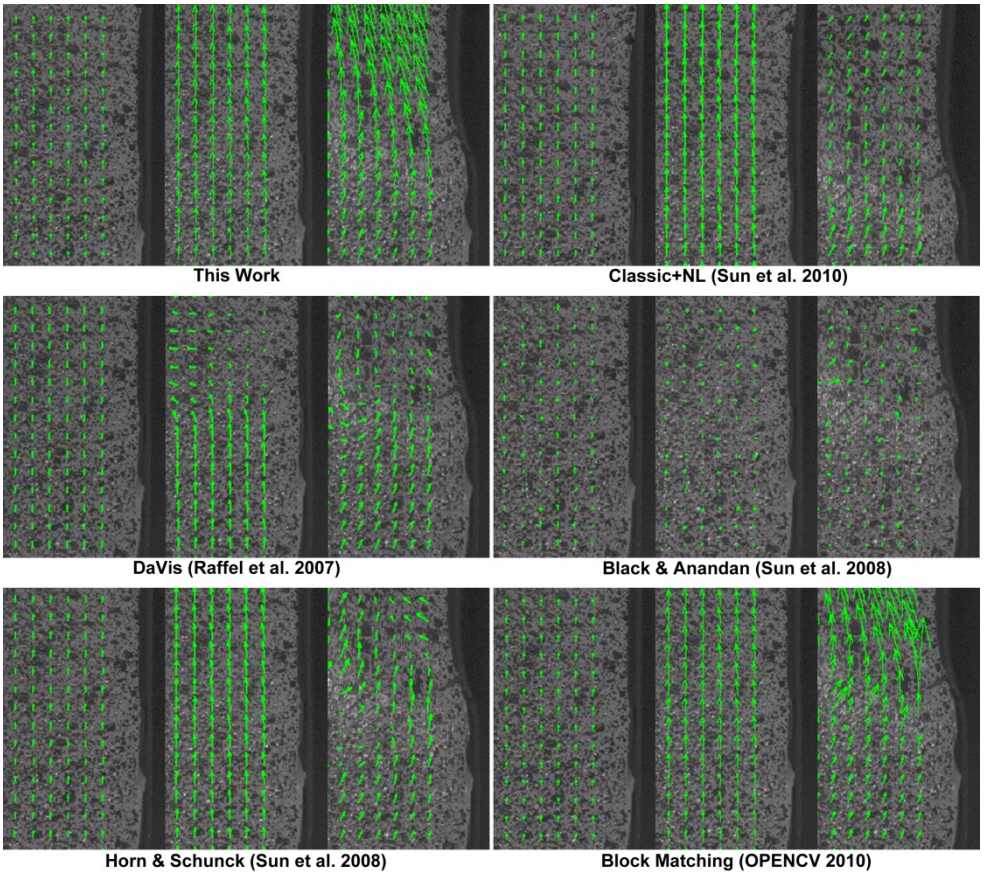

Figure 73: Visual results in the downsampled tensile test.

The same three images are shown for each technique. They represent the enlargement of an area of the material with the corresponding displacement vectors.

Table 7: Numerical results of the tensile test with a steel bar.

	High Resolution Experiment		Low Resolution Experiment	
	Avg. Error (mm)	Std.Dev.Error (mm)	Avg. Error (mm)	Std.Dev.Error (mm)
This application	0.008	0.004	0.031	0.028
Block-Matching [75]	0.046	0.036	0.171	0.326
Classic+NL [154]	0.062	0.022	1.140	2.149
Horn & Schunck [153]	1.507	1.195	1.065	1.958
DaVis [99]	1.318	2.049	1.937	2.389
Black & Anandan [153]	1.551	1.180	2.091	1.947

By analysing the obtained results, it can be seen that only the proposed technique had a good performance and accuracy during the entire sequence.

The *Classic+NL* had a good performance in the high resolution sequence but failed in the second half of the low resolution one.

The *DaVis* technique obtained only good results in the first half of the sequences. This may happen because the *DaVis* technique analyses the images in the frequency domain using Fourier Transforms to increase performance. So, as explained before, it needs to downsample the image with a pyramidal technique to analyse long displacements [99]. Therefore, when the image is downsampled, the effects of deformation are artificially increased, so that, if global characteristics of the moving object are not visible (such as the borders or the shape of the body), it may be difficult to retrieve the displacement of the surface.

The *Block-Matching* technique had a poor accuracy and obtained several anomalous vectors. This may be attributed to a poor similarity, the lack of sub-pixel precision and an inappropriate postprocessing stage.

The *Black & Anandan* and the *Horn & Schunck* algorithms had a poor performance in both experiments.

6.3.2 Real strength tests with a concrete specimen

In the second experiment, a compression test was performed on a concrete test specimen. A cubic specimen (*10×10×10 cm*) of high strength concrete was used, and similar strains can be assumed in the vertical faces of the cube when an increasing load is applied.

In these test, the concrete will show small strains until the material reaches the failure. A strain gauge is commonly used to measure these strains. A strain gauge is a device which uses electrical conductance to measure small strains with high accuracy.

Strain gauges need to be attached to the material before the test, provide measurements in a single direction and they fail when the material breaks. Furthermore, if a strain gauge is loaded beyond its design limit its performance degrades and cannot be recovered. Normally good engineering practice suggests not to stress strain gauges beyond 3000 microstrains.

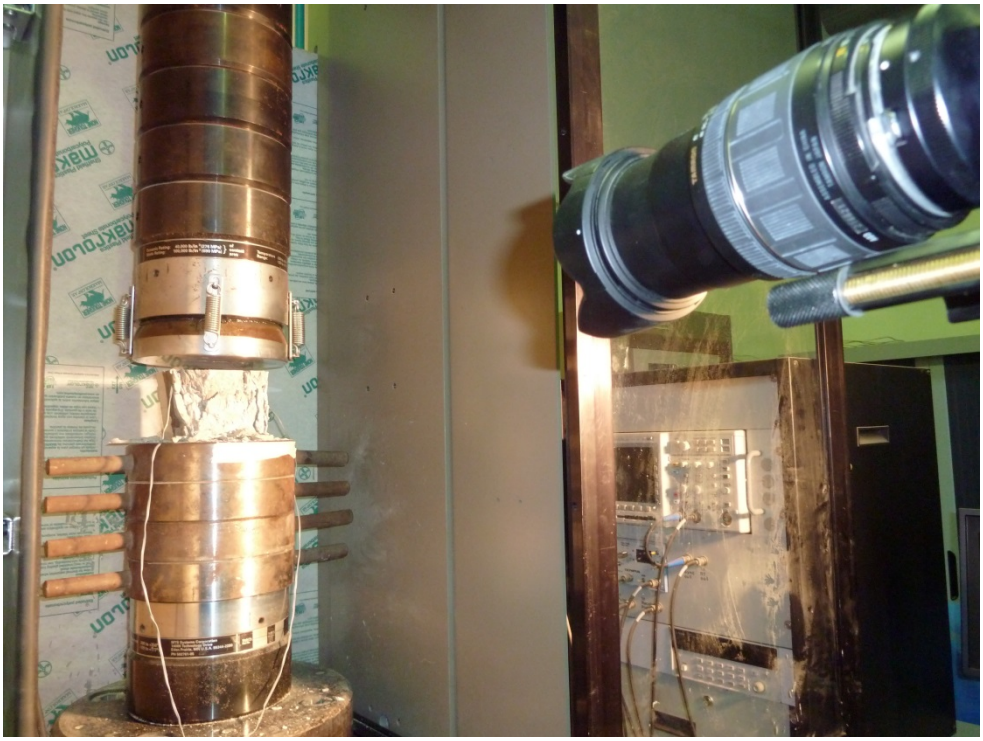

Figure 74: Execution of the compression test.

Pictures of the compression strength test with a concrete specimen performed at the Center of Technological Innovation in Construction and Civil Engineering (CITEEC).

In this test, deformation was measured in one of the vertical faces of the specimen through the proposed technique, and the obtained results were compared with those provided by two strain gauges placed in two different vertical faces of the cube. Constant loading rate was applied through the servo motor of the compression machine.

The execution of this experiment is illustrated in Figure 74 and Figure 75 illustrates how measurements were performed. The analysed sequence has a *FOV* of *100mm* and a resolution of *4* megapixels (*mm/pixel = 0.05*).

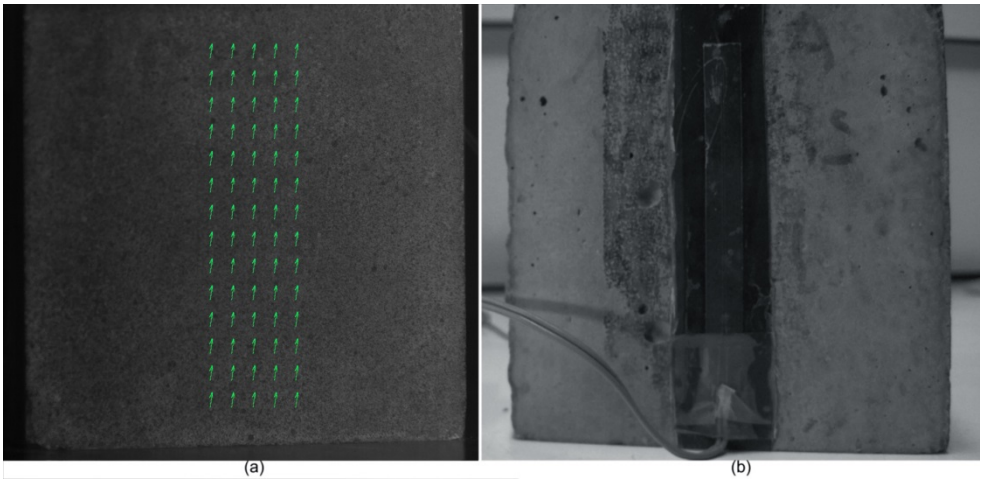

Figure 75: Visual results of the compression test.

Displacement vectors estimated during the test (a), strain gauge attached to a face of the specimen (b).

Specific longitudinal strain in microstrains is shown for each device together with the information from the compression machine in Figure 76.

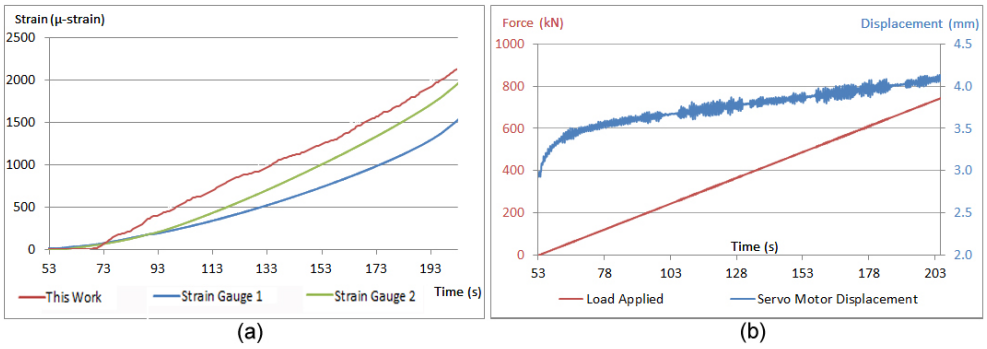

Figure 76: Results of the compression test.

Results obtained in the compression test with the concrete specimen using the proposed technique and strain gauges. Strain vs. time (a). Loading rate and displacement vs. time relationship of the stroke of the compression servo-controlled machine (b).

Analysing the obtained results it may be seen that the proposed technique produces similar results to those by strain gauges.

It can be assumed that similar or better accuracy than a strain gauge can be obtained by using a smaller *FOV* or a higher resolution camera.

6.3.3 Real strength tests with an aluminum bar

In the last experiment, a tensile test was performed with a cylindrical aluminum bar. The size of the bar was *30cm* (length) x *8mm* (diameter).

Aluminum is much more deformable than concrete, so it is not possible to use standard strain gauges to measure strains at failure. Therefore, a common contact extensometer was used to contrast the results obtained with the proposed technique.

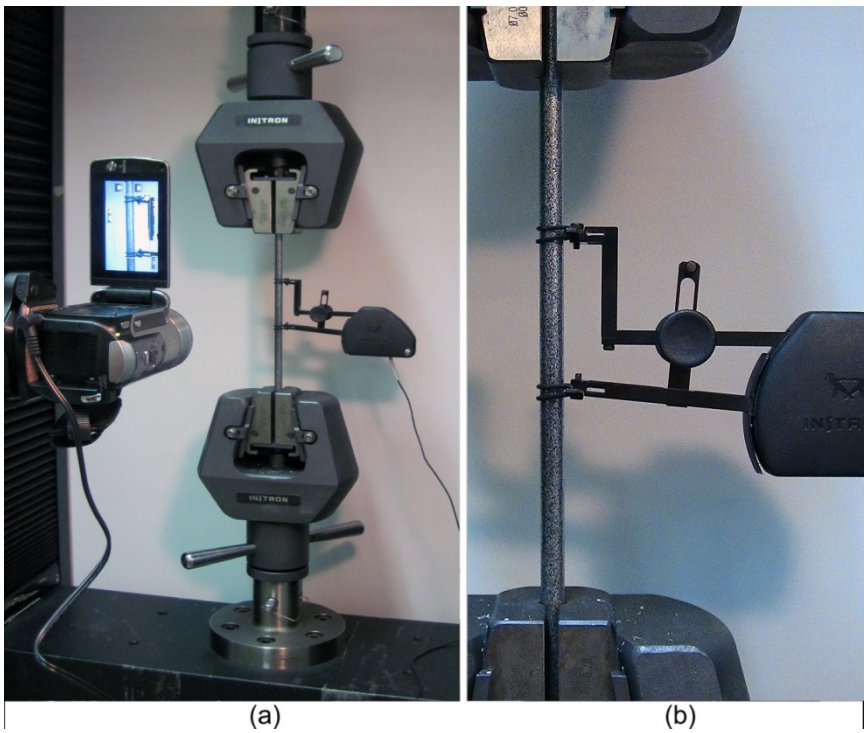

Figure 77: Execution of the second tensile test.

Pictures of the tensile strength test with an aluminum bar performed at the Civil Engineering School (A Coruña University).

The execution of this experiment is illustrated in Figure 77 and Figure 78 illustrates how measurements were performed. The analysed sequence has a *FOV* of *100mm* and a full HD resolution (*mm/pixel = 0.05*).

In this case it was applied a constant rate of stroke displacement, with an intermediate section in which the displacement was held fixed.

Figure 79 shows the specific strain provided by the proposed technique and by the extensometer.

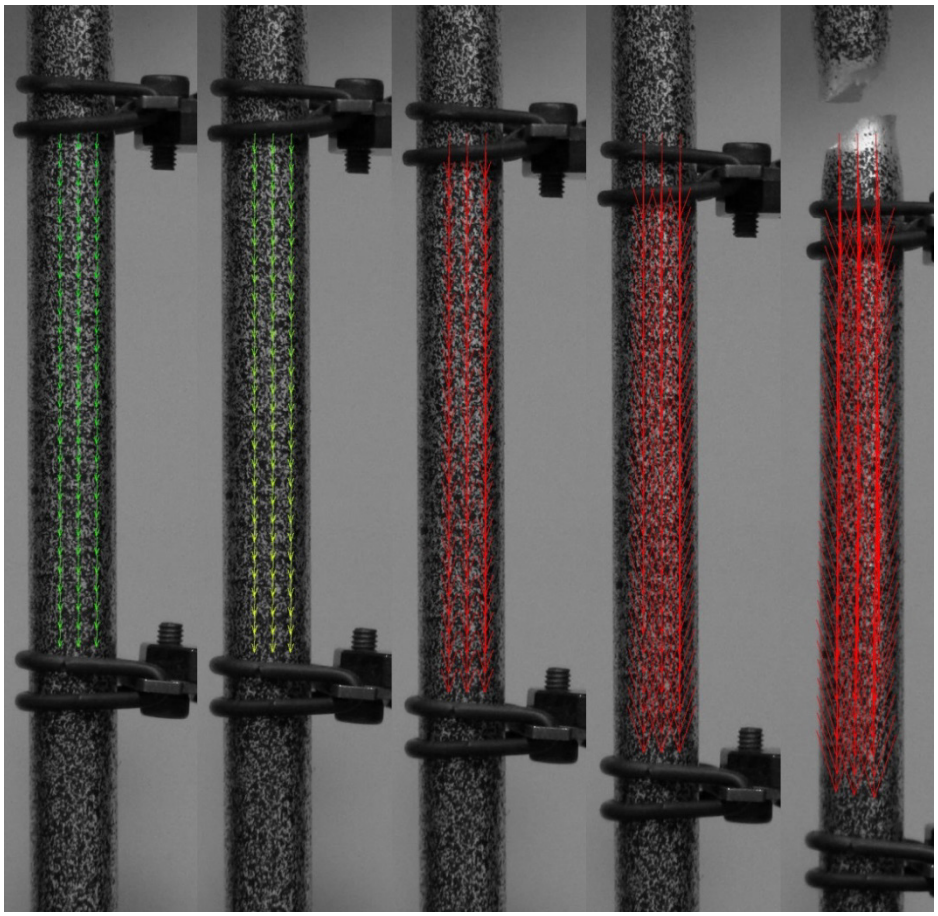

Figure 78: Visual results of the second tensile test.

Displacement vectors estimated during the test and extensometer attached to the material.

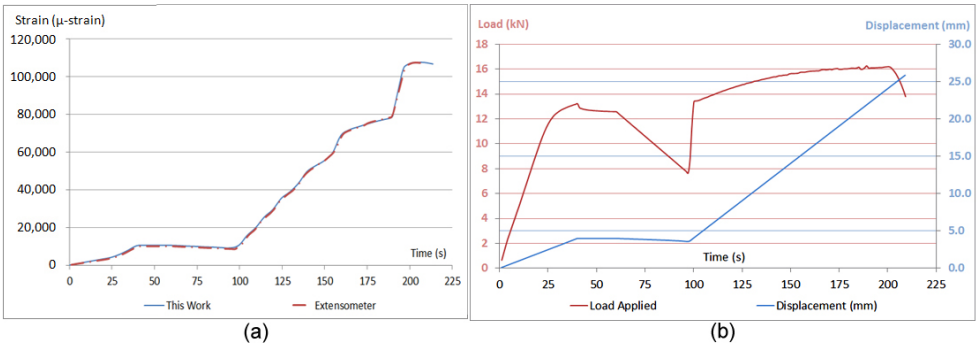

Figure 79: Results of the second tensile test.

Specific strain (in microstrains) provided by the proposed technique and by the extensometer (a). Load and displacement vs. time curves during the test (b).

Analysing the obtained results it may be seen that the proposed technique produces virtually identical results to those by extensometer.

7 Conclusions

This research introduces a new technique to analyse the strains using block adjustment for measuring displacements.

The proposed technique is a general solution for the 2D deformable case. It can be used with different techniques and configurations according to the requirements of the problem and uses different deformation models with different degrees of freedom which can be imposed in different stages of the algorithm, so the behavior of the analysed body can be obtained by using a regular grid of blocks and a local deformation model or by analyzing an arbitrary set of points and a global model.

Furthermore, different similarity metrics, subpixel algorithms and postprocessing filters can be easily used to analyse different scenarios.

The proposed technique has been compared with different computer vision techniques to measure displacements in extensive assays, obtaining the best accuracy in terms of error in the test using standard benchmarking images.

The proposed technique has also been compared with traditional instrumentation such as strain gauges and contact extensometers, obtaining similar results.

Furthermore, this technique has several advantages: first, the cost is lower, and moreover the measurement process is more flexible, because it is possible to measure any strain range without having to select previously any region of interest.

In addition, unlike any traditional technique, the proposed one retrieves the complete displacement field of the surface, providing information about the global behaviour of the material being tested.

Some of the finds of this applications were published in different media, the main of these contributions are listed below:

Freire, J. A. Seoane, A. Rodriguez, C. Ruiz Romero, G. López-Camos, and J. Dorado, "A Block Matching based technique for the analysis of 2D gel images," Studies in health technology and informatics, vol. 160, pp. 1282-1286, 2010.

A. Rodriguez, C. Fernandez-Lozano, J.-A. Seoane, J. Rabuñal, and J. Dorado, "Motion Estimation in Real Deformation Processes Based on Block-Matching Techniques," presented at the IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), 2011.

A. Rodriguez, C. Fernandez-Lozano, J.-A. Seoane, J. Rabuñal, and J. Dorado, "Analysis of Deformation Processes Using Block-Matching Techniques,"

presented at the International Conference on Computer Vision Theory and Applications, Roma, Italia, 2012.

A. Rodriguez, J. Rabuñal, J. L. Perez, and F. Martinez-Abella, "Optical Analysis of Strength Tests Based on Block-Matching Techniques," Computer-Aided Civil and Infrastructure Engineering, vol. 27, pp. 573–593, 2012.

A. Rodríguez , C. Fernández-Lozano, and J. Rabuñal, "Spot matching in 2D electrophoresis experiments," International journal of Data Analysis and Strategies, vol. 5, pp. 198-213, 2013.

A. Rodriguez, C. Fernández-Lozano, J. Dorado, and J. Rabuñal, "2-D Gel Electrophoresis Image registration using Block-Matching Techniques and Deformation Models," Analytical Biochemistry, 2014 (Accepted, Publication Pending).

A. Rodriguez and J. Rabuñal, "Measuring Deformations in Biomedicine with Block-Matching Techniques: A Survey," 2014. (Being Processed).

The most remarkable of these published applications is the use of the proposed technique in the biomedical problem of registering 2D electrophoresis gels, which consists in finding the correspondence in two images representing 2 dimensional patterns of proteins. Where each protein appears as a dark spot and assuming that one of the images can be mapped into the other applying a specific deformation. An example of the results obtained in this scenario is shown in Figure 80.

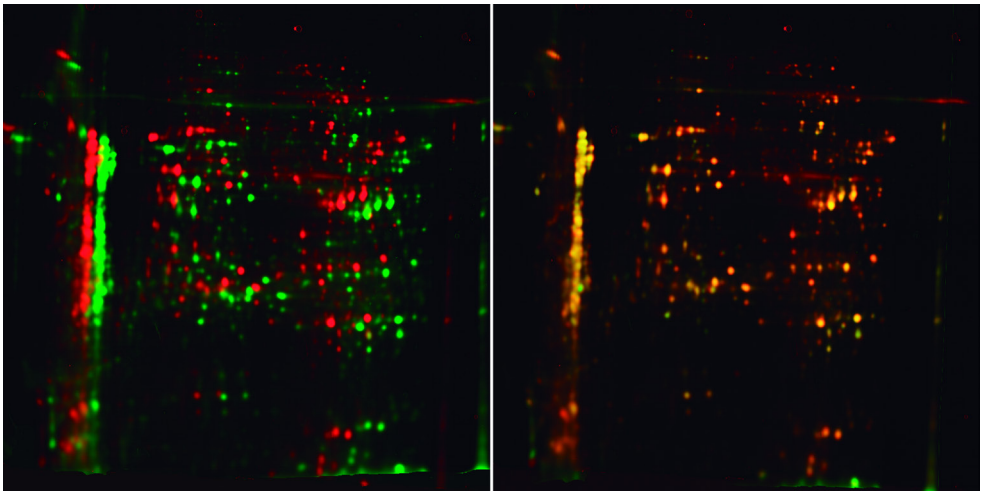

Figure 80: Example of 2D electrophoresis gel registration.

(Left): Reference image, colored in red, superimposed to the original test image, colored in green. (Right): Reference image, colored in red, superimposed to the registered one, colored in green. Overlapped spots can be seen in yellow.

Some of the further research which may be conducted will be the adaptation of the technique to obtain 3D measurements with a multicamera system, and the analysis of different materials and different experimental conditions.

I realized then that we could not hold our own with the white men... They would change the rivers and mountains if they did not suit them.

Heinmot Tooyalaket

VI. Application in Fish Tracking

Vertical slot fishways are hydraulic structures which allow the upstream migration of fish through obstructions in rivers. The appropriate design of a vertical slot fishway depends on interplay between hydraulic and biological variables, since the hydrodynamic properties of the fishway must match the requirements of the fish species for which it is intended.

One of the main difficulties associated with studies of real fish behavior in fishway models is the fact that the existing mechanisms to measure the behavior of the fish in these assays, such as direct observation or placement of sensors on the specimens, are impractical or unduly affect the animal behavior.

This thesis proposes a new procedure for measuring the behavior of the fish. The proposed technique uses different *CV* techniques to analyze images obtained from the assays by means of a camera system designed for fishway integration. It is expected that this technique will provide detailed information about the fish behavior and will help to improve fish passage devices, which is currently a subject of interest within the area of Civil Engineering.

A series of assays have been performed in order to validate this new approach in a full-scale fishway model and with living fishes. We have obtained very promising results that allow reconstructing correctly the movements of the fish within the fishway.

1 Introduction

The construction of water resources management works, such as dams, weirs, water diversions and other barriers leads to significant changes in the river ecosystem. These structures constitute a physical barrier to fish natural movements, which negatively impacts their populations. In fact, this interruption of free passage has been identified as the main reason for the extinction or the depletion of numerous species in many rivers [161].

Out of the various solutions employed to restore fish passage, some of the most versatile are known as vertical slot fishways. This type of fishway is basically a channel divided into several pools separated by slots. Its main advantage lies in its ability to handle large variations in water levels, since the velocity and turbulence fields in the pools are independent of the discharge. Moreover, it allows fish to swim at their preferred depth and to rest in low-velocity areas, in contrast to other types of fishways.

An effective vertical slot fishway must allow fish to enter, pass through, and exit safely with minimum cost to the fish in time and energy. Thus, biological requirements such as fish preferences should drive design and construction criteria for this type of structures. However, while some authors have characterized the flow in vertical slot fishways [162-164] and others have studied fish swimming performance [165, 166]. Besides, very few works have studied the interaction between the biological and physical processes that are involved in swimming upstream a vertical slot fishway [167].

Consequently, the knowledge of fish behavior when confronted with this type of structures is limited and these biological requirements usually rely on the designer's experience, rather than on rational approaches.

In order to address this deficit, it is necessary to complete the fishway design methodology with results from experimental assays with living fish. In these tests, fish are introduced into full-scale fishway models such as the shown in Figure 81, and their movements and behavior are analyzed. In the tests described in this thesis, the passage success (the proportion of individuals that passes through the fishway) is evaluated and the fish effort is measured by means of blood tests. Nevertheless, these techniques should be used in combination with another that permits detailed characterization of the animal behavior during the assay, determining parameters such as: resting areas, resting times, fish velocities and accelerations, times spent for full ascent and in each pool, etc. Subsequently, these parameters can be linked to the hydraulic data and the results of the blood tests in order to define fish preferences and requirements. The results can be immediately applied to the design methodology of these devices, establishing such hydraulic conditions in the fishways that the target swimmers can negotiate it. Consequently, the vertical slot fishways designed by Civil Engineering experts with this new methodology will be more efficient and fish passage will be enhanced.

Figure 81: Fishway model.

(Center for Studies and Experimentation of Public Works, Madrid, Spain).Indoor full-scale (1:1 scale) vertical slot fishway model used in this study. Photo taken at CEDEX

However, the monitoring of the fish on the fishway during the assay with the current methods gives rise to many difficulties:

- *Direct observation*: This technique is impractical because of the difficulty of observation due to the water turbulence and the limited validity of the information collected.
- *Placement of sensors on the specimens:* This technique is based on the placement of antennas in key positions in order to record the pass of the fish, which will be equipped with a transmitter [168]. Although this technique may be a good alternative when the tests are carried out with a large number of individuals, it provides no information on the full fish trajectory. Besides, the sensors placed externally tend to fall off, and the surgically-implanted sensors affect significantly the animal behavior.

Therefore, it is necessary to develop a new technique to measure the behavior of the fish within the fishway, in a less intrusive way and capable of obtaining more accurate information than direct observation and placement of sensors techniques. To this end, an approach based on optical or acoustic monitoring is the best alternative. Some early examples of these applications are the use of acoustic transmitters and a video camera for observing the behavior of various species [169], or the utilization of acoustic scanners for monitoring fish stocks [170].

More recently, different *CV* techniques for the study of the fish behavior have been used; these works use techniques such as stereo vision [171], background models [172], shape priors [173], local thresholding [174], moving average algorithms [175], particle image velocimetry techniques [156], pattern classifiers applied to the changes measured in the background [176] or Artificial Neural Networks (*ANN*) [38]. Finally, some techniques based on infrared imaging [177] or *LIDAR* (Light Detection and Ranging or Laser Imaging Detection and Ranging) technologies [178] have also been studied.

However, it should be noted that most of the published techniques have been carried out in calm water conditions and with controlled light sources, and therefore are not suitable for their use inside a fishway. Additionally, some of these techniques use marks or light sources, which may influence fish behavior, while others employ special and expensive sensors, which may be only used in certain points of the structure.

Consequently, since none of the current techniques would be appropriate in the context of this research, a new technique is proposed to study fish behavior in vertical slot fishways. It gathers information at every instant on the position, velocity and acceleration of the fish. To this end, images obtained through a

network of video cameras are analyzed using *CV* techniques and artificial neural networks.

As a part of the technique proposed, a new algorithm to extract the mass center of the fishes in the images segmented by the neural network is presented, as well as a procedure to integrate the different cameras used in the assays.

2 Proposed Technique

The proposed technique calculates the position of the fish on the fishway at every instant from the recorded images of the assay.

To this end, a camera system equipped with fisheye lenses that provide a 180º viewing angle has been installed. The cameras were placed in an overhead perspective and partially submerged so that the entire fishway is covered and turbulence and surface reflections are avoided.

Figure 82 show real images taken during the assays, showing the location of the cameras with water protection structure, illustrating the operation of the camera and it shows the visual angle of the camera slightly reduced by the protection structure.

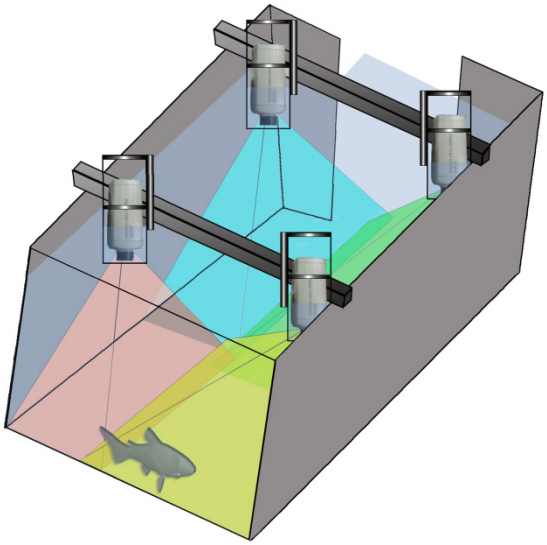

Figure 82: Recording conditions

Therefore, a total of 28 cameras have been used, 4 per pool. The cameras have been integrated into the monitoring and data acquisition system so that each camera has been connected to a hard disk and the recording of all disks has been synchronized. In addition, each hard disk has been connected to a central computer and a monitor in order to facilitate the control of the process. The data acquisition system is shown in the schematic of Figure 83.

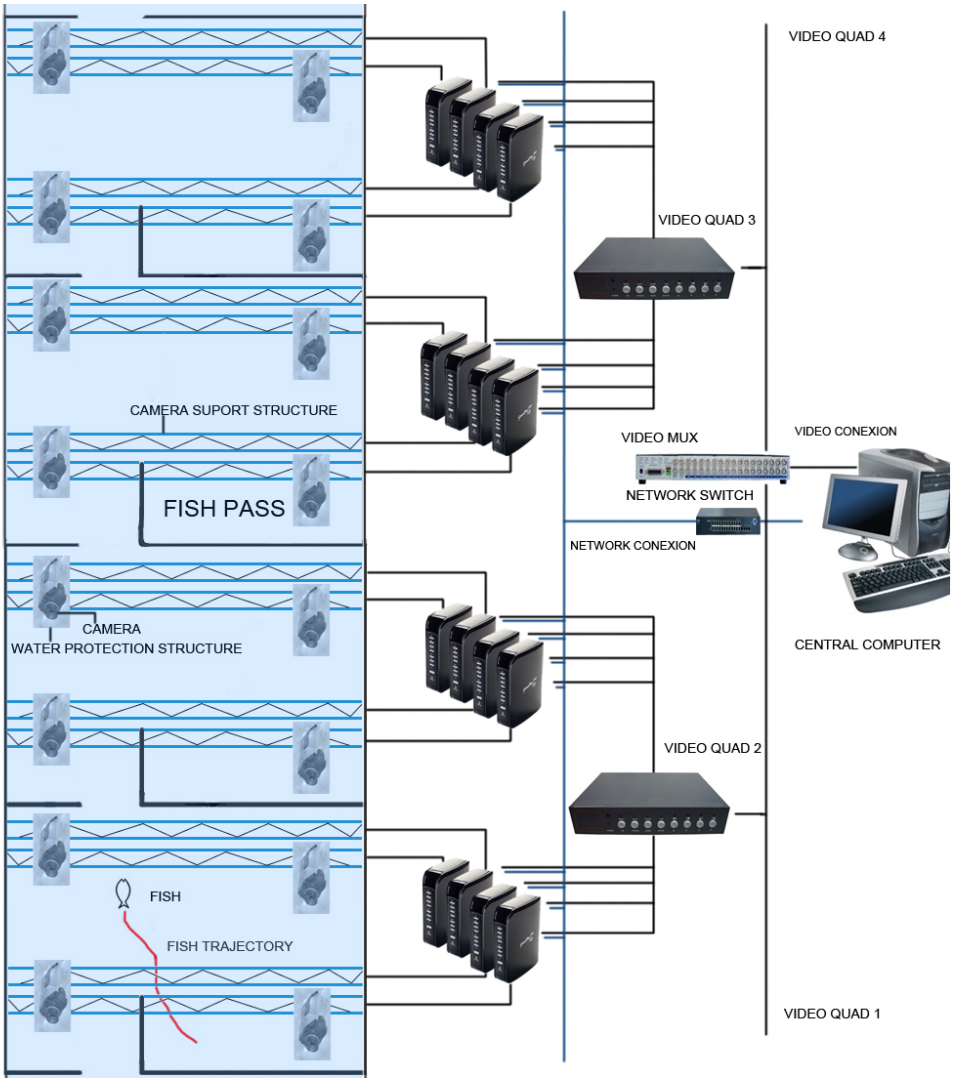

Figure 83: Overview of the data acquisition system.

From the images taken by the data acquisition system, the position of the fish on the fishway will be obtained at every instant, allowing us to calculate the velocity and acceleration of the fish. The overall performance of the technique can be summarized as follows:

1. Camera calibration: The image distortion is eliminated and a projective model is designed to integrate measurements from the different cameras into a common coordinate space.
2. Segmentation: The image background is separated from the regions where the fish can be found.
3. Representation and Interpretation: The detected objects are translated into a descriptive representation which can be operated with. A first filtering step is done at this point.
4. Tracking: The detected objects are used as the input of a filter extended from the Kalman model. This is an algorithm that uses a motion model and works in a similar way to a Bayesian filter. It allows to delete isolated noise detections, to separate different fishes, to estimate hidden positions of the fishes and to predict next positions of their trajectories.

2.1 Camera Calibration

The first step of the algorithm is to calculate the transformation from coordinates in a particular camera to real coordinates in the fishway. This calculation is performed in two steps: first, the parameters to correct and scale the image are obtained for each camera and second, the transformation of each camera into a common coordinate system is calculated.

For the first step, we use the pin-hole projective model has been used as described by Zhang [27] with the distortion model proposed by Weng et al. [74] and shown in Figure 84.

The model was solved with the standard calibration method proposed originally by Zhang [27]. This technique was described with detail in section IV.3.3.

The first step of the algorithm is to calculate the transformation from coordinates in a particular camera to real coordinates in the fishway. This calculation is performed in two steps: first, the parameters to correct and scale the image are obtained for each camera and second, the transformation of each camera into a common coordinate system is calculated.

For the first step, we use the pin-hole projective model [27] which describes how a point from the real space is projected into the image plane as shown in Figure 84. Pin-hole equations can be written as follows:

$$\begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = M \times \begin{bmatrix} x_c \\ y_c \\ 1 \end{bmatrix} \qquad M = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 0 \end{bmatrix} \quad (55)$$

Where (x_i,y_i) are the coordinates of the point in the image and (x_c,y_c) are the projected coordinates in the space. M is called transformation matrix and is defined by the focal length (f_x, f_y) and the position of the point which is projected through the center of the lens or optical center (c_x,c_y).

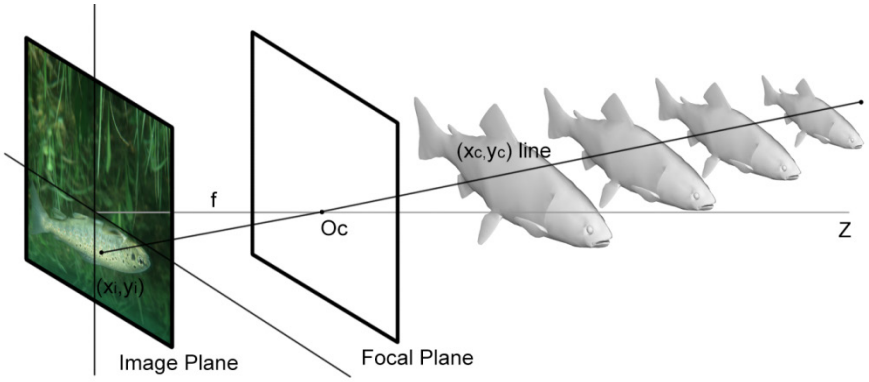

Figure 84: Projection of a fish in a pin-hole camera model.

The point (x_i,y_i) represent a pixel of a fish in a digital image, and (x_c,y_c) are the corresponding projected coordinates in the space. The image is projected through the center of the lens or optical center *Oc*.

Additionally, the refraction of light in the water must be considered. Thus, cameras should be calibrated underwater or refraction should be modeled with an additional transformation. In this case, an affine model is employed to perform this task.

In the second process, the measurements obtained by each camera are connected to a common coordinate system, the entire pool being covered. Thus, applying an equation of general projective geometry, the transformation M_2 between the global coordinate space *c'* and the coordinate space *c*, obtained from a concrete camera, can be expressed as follows:

$$\begin{bmatrix} x_{c'} \\ y_{c'} \\ 1 \end{bmatrix} = M_2 \times \begin{bmatrix} x_c \\ y_c \\ 1 \end{bmatrix} \qquad M_2 = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & 1 \end{bmatrix} \tag{56}$$

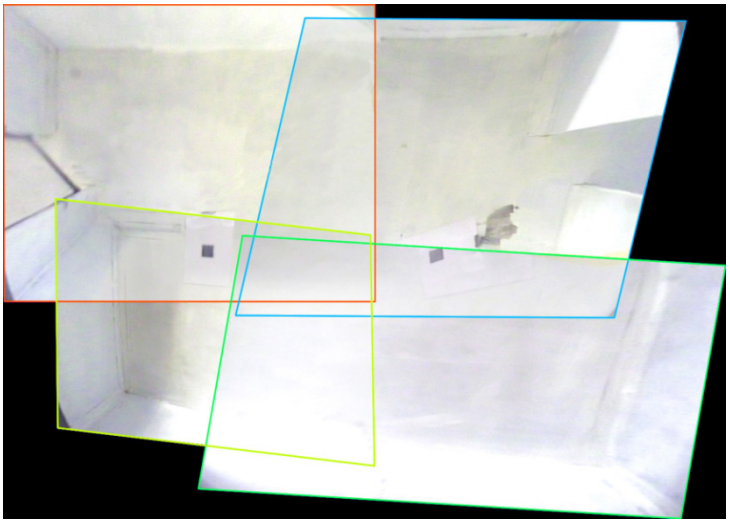

Figure 85: Camera overlapping vision fields.

The image shows a pool of the fishway formed with the projected images from the four cameras which are covering the pool. Each polygon represents the field of each different camera.

To solve the equation above, a number of visual marks have been placed in the areas where the vision field of the cameras overlaps (Figure 85).

2.2 Segmentation

Image segmentation is the process of partitioning an image into multiple parts, in this case, with the aim of separating the fish from the background. Every segmentation algorithm addresses two problems, the criteria for a good partition and the method for achieving an efficient one [179]. In statistics, this problem is known as cluster analysis and is a widely studied area with hundreds of different algorithms [13].

Therefore, it is first necessary to find a variable or group or variables (features) which allow a robust separation of the fish from the background and then to choose a technique to classify the image according to the selected variables.

In the state of the art, most common criteria to detect fishes in images are based on color features and *a priori* knowledge of the background. However, these techniques do not perform well in underwater images, even for calm water and high quality images, due to the low levels of contrast [176]. Besides, acquired images in this study will be characterized by extreme luminosity changes and huge noise levels, being texture and color information useless.

Taking this into account, different techniques are considered in this work. These techniques are detailed in the System's performance section and provide a comparative framework. They are based on the discontinuities in the intensity of the image (edge-based classification) or in the intensity similarity of pixels (region-based classification). Due to the necessity to operate the systems by non-experts, only non-supervised techniques have been considered and, given the huge

amount of images to be analysed, computational complexity was decided to be a critical factor.

One of these techniques, previously developed in [38], consists of a *SOM* (Self-Organizing Map) neural network [180]. The *SOM* model is aimed at establishing a correlation between the numerical patterns supplied as an input and a two-dimensional output space (topological map). This characteristic can be applied to image segmentation, and *SOM* networks have been widely used in the image analysis field in different works [136, 181-184]. Although promising results were obtained in this early work, the *SOM* approach is slow compared with more straightforward techniques and it depends on the training patterns selected.

In this work, a combination of two simple techniques is selected, together with an image preprocessing procedure and a dynamic background modeling, to overcome the limitations of the *SOM* approach.

The first selected technique a modern implementation of the *Canny* edge detector, which resulted less noisy than other edge-finding methods such as the *Sobel* or the *Prewitt* operators. With this technique, edges corresponding to the frontiers between fishes and background are obtained by means of 4 directional filters to find the horizontal, vertical and diagonal discontinuities in the derivatives of the images [185] .

In order to reduce the false positives obtained, the objects detected by the region analysis are filtered using a second segmentation technique. Thus, only the detected objects overlapped in a *95%* per cent with those obtained with the second technique are considered. This second technique is the *Otsu* method, which performs a region classification by automatically thresholding the image histogram. This technique is a fast and efficient technique to separate dark objects in a white background [186].

So as to introduce knowledge about the background in the method, a dynamic background is calculated forming a synthetic image and is used to normalize the image. The synthetic background is calculated according to the following equation:

$$BI_F(x,y) = 0.5 \times BI_P(x,y) + 0.5 \times I_F(x,y) \quad (57)$$

Where, BI_F is the new background image, I_F is the current image, corresponding to the frame F of the video, and BI_p is the previous background image, calculated in the frame P. The new frame is updated separately for each quarter of the image if $F\text{-}P > 30$, and no objects were detected in the quarter of the image to be updated. In $F=1$ a static image with no fish is used as background.

In order to enhance the image quality, the images are preprocessed using a standard contrast-limited adaptive histogram equalization technique. Besides, the borders where the waterproof cases of the cameras produced a black region (without information) were masked.

2.3 Representation and Interpretation

Once the image segmentation is obtained, the image is divided in different regions representing background and possible fishes. In this stage of the process, it is possible to use a higher level processing, using knowledge extracted from the characteristics of real fish, to perform the interpretation of the segmented image. To this end, the objects detected in the previous stage are translated into convenient descriptors, which can be used by the computer to perform different operations: its area, its centroid (calculated as the average position of the body pixels) and the minimum ellipse containing the body.

Subsequently, an algorithm has been built to classify each detected body into fish or non-fish categories, according to the values of these properties. The operation of this algorithm is schematized in Figure 86, and Figure 87 shows the obtained results when applying image interpretation to the algorithm.

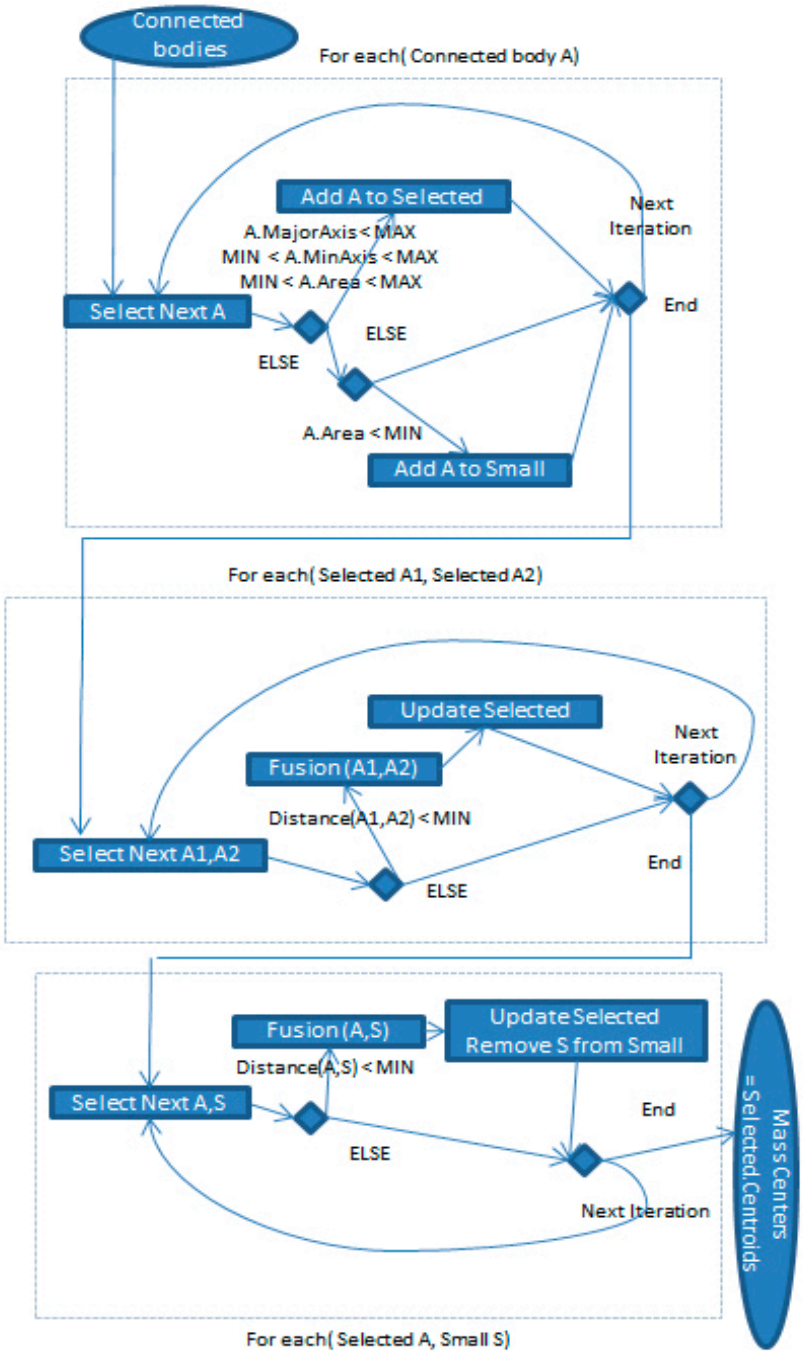

Figure 86: Diagram of the algorithm used to interpret the segmented image.

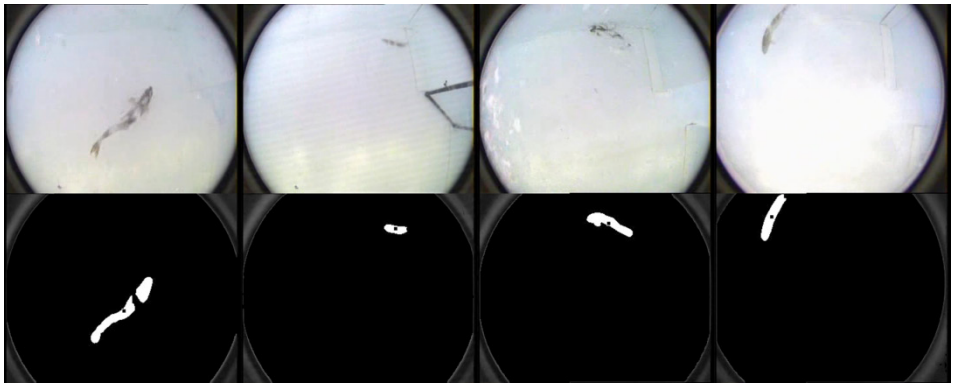

Figure 87: Obtained results after segmentation and interpretation steps.

Original images are shown at the top of the figure and their corresponding segmentation is shown at the bottom. Each detected body is marked in white and each fish center is marked with a small dark square inside the body. Note that some bodies have been interpreted as parts of the same fish.

2.4 Tracking

Tracking is the problem of generating an inference about the motion of one or more objects from a sequence of images. It can be solved using several approaches, including motion estimation and feature matching techniques. Some of the most important approaches consider a statistical point of view and formulate the problem as a prediction correction process based on the Bayesian theory.

A well-known technique in this group is the Kalman Filter [53, 187-189]. It addresses the problem of estimating the state $x \epsilon R^n$ of a discrete-time controlled process that is governed by the linear stochastic difference equation, expressed as follows:

$$x(t+1) = Ax(t) + w(t) \quad (58)$$

Where A is a nxn matrix called state transition matrix, which relates the state of the system at the previous time step to the state at the current step, and w represents the process noise, which is assumed normally distributed with mean 0.

For the state transition matrix, we consider the equations of two dimensional motion assuming a constant acceleration between time steps:

$$\begin{aligned} x_{t+1} &= x_t + v_{x,t} + \frac{1}{2}a_x t^2 \\ y_{t+1} &= y_t + v_{y,t} + \frac{1}{2}a_y t^2 \\ v_{x,t+1} &= v_{x,t} + a_x t \\ v_{y,t+1} &= v_{y,t} + a_y t \end{aligned} \quad (59)$$

Where (x, y) is the fish position, (v_x, v_y) is the velocity and (a_x, a_y) is the acceleration, which is assumed constant in the time interval t (equal to the frequency of the image acquisition, i.e., *0.04* s).

We consider also an observation model described by the following equation:

$$z(t) = Hx(t) + v(t) \quad (60)$$

Where $z \in R^m$ represents the measurement, H is a mxn matrix called observation matrix and v is a measurement error assumed independent of w and normally distributed with mean 0.

After the calibration process, it is assumed that observed positions in the image will provide real world positions. In addition, the acceleration term can be assumed to have a zero mean, so the model equations can be represented as follows:

$$\begin{bmatrix} x_{t+1} \\ y_{t+1} \\ v_{x,t+1} \\ v_{y,t+1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & t & 0 \\ 0 & 1 & 0 & t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_t \\ y_t \\ v_{x,t} \\ v_{y,t} \end{bmatrix} + w(t)$$
$$\begin{bmatrix} z_x \\ z_y \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_t \\ y_t \\ v_{x,t} \\ v_{y,t} \end{bmatrix} + v(t) \qquad (61)$$

Using this model, the Kalman filter works in a two-step recursive process: In the prediction step, the Kalman filter estimates the new state, along with their uncertainties. Once the outcome of the next measurement (corrupted with noise) is observed, these estimates are updated using a weighted average in the second step. The higher weight is given to the estimates with higher uncertainty. The algorithm can, therefore, run in real time using only the present input measurements and the previously calculated state.

In the present work, the implementation of the Kalman filter was performed according to [190], obtaining an empirical estimate of the measurement error and the process noise covariances.

The filter is designed to track multiple objects, which are referred to as fishes or tracks. The essential problem which is solved at this point is the assignment of detections to fishes.

In order to associate a detection to a track, a cost is assigned to every possible pair fish-detection, understood as the probability for that detection to correspond to the current position of the fish. It is calculated using the distance from the detected position to the predicted position and to the last confirmed position for the fish. To this end, the minimum of the Euclidian distances is selected as a cost metric.

Therefore, every detection is assigned to the track with the lower cost, provided that it is lower than a selected value, and each track can only be assigned to one detection. When a new detection is assigned to a fish, the predicted position for that instant is confirmed and corrected. Detections which remain unassigned to any existing fish are assumed to belong to new tracks. Additionally, if a fish remains unassigned for too long, its track is closed, so no new assignments can be made to that fish. Fishes without enough detections are assumed to be noise and deleted. The operation of the assignment algorithm is described in the schematic of

Figure 88 and the results obtained in a situation with two fishes are shown in Figure 89.

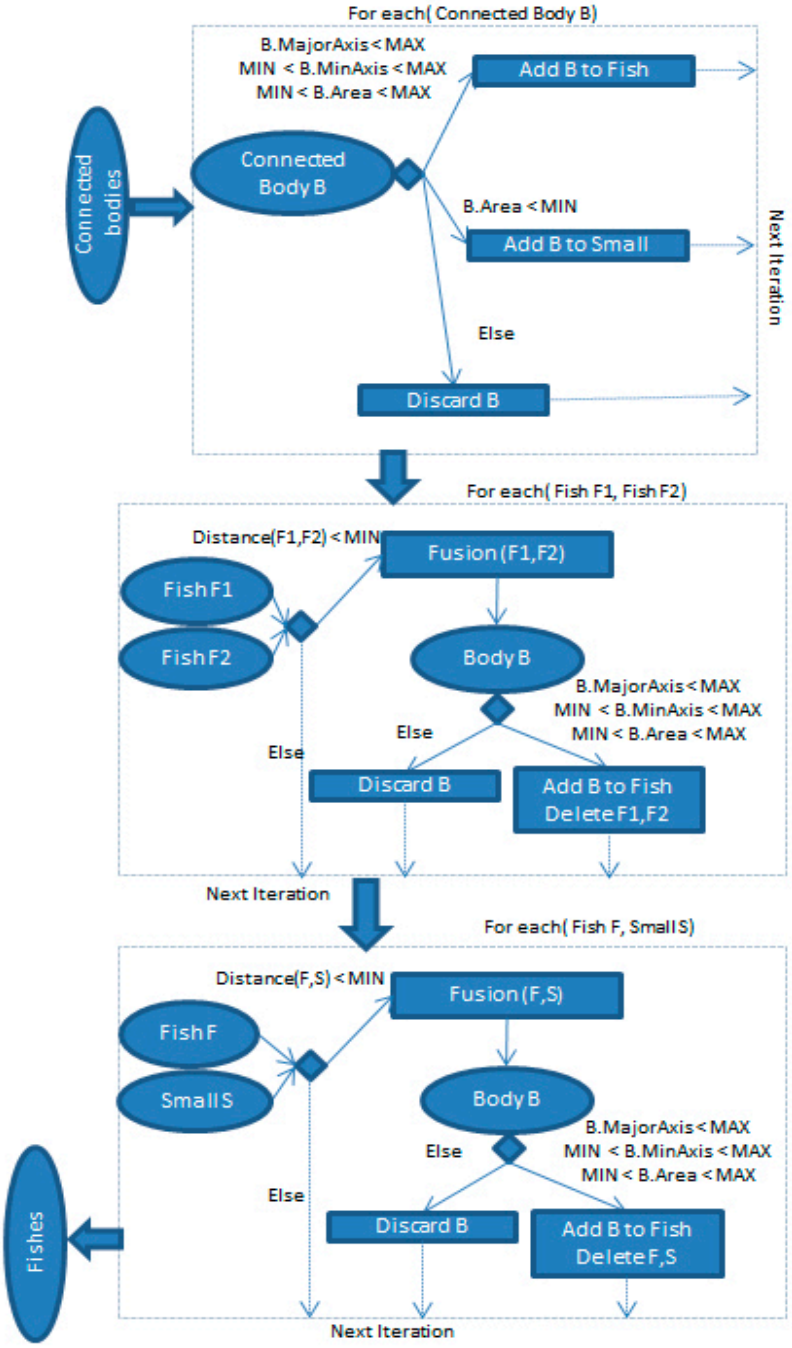

Figure 88: Diagram of the algorithm used to assign detections to fishes.

In conclusion, this technique does not only obtain trajectories from detections, but also allows filtering some of the false positives of the system and estimating the fish position when it is not detected in the images.

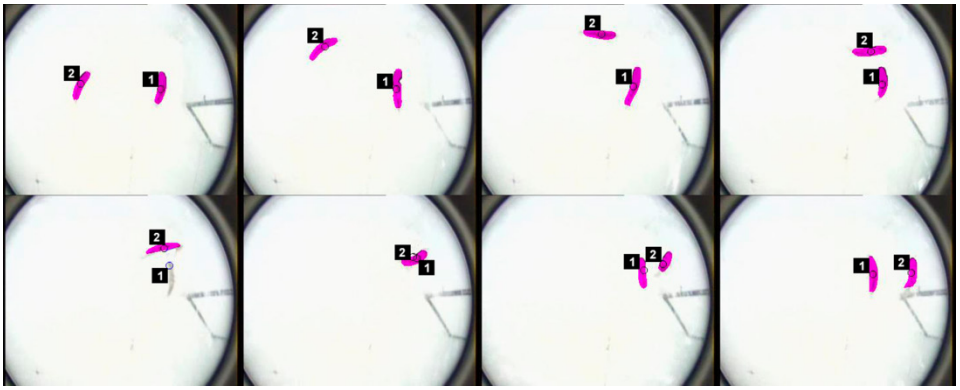

Figure 89: Obtained results with Kalman filtering in a sequence of images.

Image order is from left to right and from the top to the bottom. Fish detections are shown as dark areas superimposed to the images. Black circles and labels indicate tracked fishes. Note that the position of the two fishes is calculated even if they are not detected in the image.

2.5 Filtering

The result of the process so far is the vector of positions of every detected fish along time, representing its full trajectory on the fishway. However, it is expected that these results will still show certain undesirable phenomena, caused by the small variability of the calculated position of the centroid, the parts of fish which are hidden by bubbles and the errors in perspective and alignment of planes, when the fish moves from a field of view of a camera to another.

In order to solve these problems, and to remove some of the noise still present in the results, a complex filtering process is required. It is performed in two steps. First, the relative position of the cameras is taken into account to solve differences between simultaneous observations. Thus, when the fish is detected

simultaneously by two or more cameras, its position is the average of all the observed positions. Secondly, a moving average filtering process is applied [191] and outliers are detected by thresholding the distance of each original position with the filtered one.

Therefore, while normal detections are simply replaced by their filtered ones, the outliers are replaced by the average of the previous and next confirmed detections. This implies that predicted positions near outliers are no longer valid, and hence they are replaced using interpolation techniques.

2.6 Data Analysis

As a result of the previous process, the fish position on the fishway over time is obtained, which leads to the definition of a position vector, as follows:

$$\left[\left(x_{t_0}, y_{t_0}\right), \left(x_{t_1}, y_{t_1}\right), ..., \left(x_{t_i}, y_{t_i}\right), ..., \left(x_{t_N}, y_{t_N}\right)\right]_{\min(\Delta t) = 0.04s} \quad (62)$$

Where x_{tN} is the x coordinate of the fish in the global coordinate system in time t_N and y_{tN} is the y coordinate of the fish in the global coordinate system in time t_N.

From the fish position vector, the observed instantaneous velocities are obtained:

$$\overline{v}_{obs} = \left(\frac{X_{ti} - X_{ti-1}}{t_i - t_{i-1}}, \frac{Y_{ti} - Y_{ti-1}}{t_i - t_{i-1}} \right) \quad (63)$$

Where $\bar{v}_{obs}$ is the observed fish velocity vector.

However, the observed velocities are not really those which quantify the real effort made by fishes to swim. In order to calculate the actual swimming velocities, the water velocity in the pools must be taken into account, as follows:

$$\overline{v}_{swim} = \overline{v}_{obs} - \overline{v}_{flow} \qquad (64)$$

Where $\bar{v}_{swim}$ is the fish swimming speed vector and $\bar{v}_{flow}$ is the flow velocity vector.

The water velocity in the fishway can be evaluated by means of experimental studies [162-164] or numerical models [192-194]. In this case, the velocity field in the pools is computed with a numerical model based on the 2D depth averaged shallow water equations. The experimental validation of this model in 16 different fishway designs, as well as a detailed description of the model equations, can be found in [195]. Figure 90 shows the computed velocity field in two consecutive pools of the fishway model.

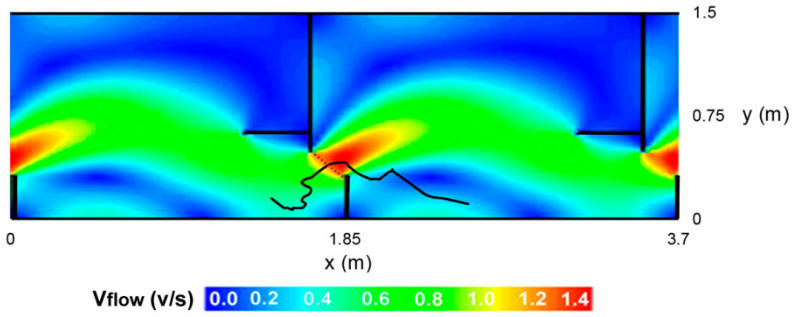

Figure 90**:** Example of fish ascent trajectory.

Once the fish swimming velocities are known, their accelerations $\bar{a}_i$ are calculated according to the following expression:

$$\overline{a}_i = \left(\frac{v_{swimx,ti} - v_{swimx,ti-1}}{t_i - t_{i-1}}, \frac{v_{swimy,ti} - v_{swimy,ti-1}}{t_i - t_{i-1}} \right) \qquad (65)$$

Where v_{swimx} and v_{swimy} are the x-component and the y-component of the fish swimming velocity.

In addition to these indicators, further information regarding fish behavior can be obtained, such as ascending and resting times, total distance covered or preferential areas for rest. Although further research is needed, the analysis of these parameters can contribute to the definition of key factors in fish passage through these devices.

3 Experimental Results

3.1 System's Performance

In order to measure and compare the performance of the system, several experiments were performed in a full-scale vertical slot fishway model located in the CEDEX laboratory. From these experiments, a dataset was created to apply the proposed technique and other well-known methods. The dataset is composed of 135 videos with a total of 11,028 images, corresponding to different cameras, pools and fishway regions, as well as different daylight conditions and fish species.

These videos were manually labeled by experts who marked in the images the fish center positions, and this information was used as ground-truth data to evaluate the techniques. In order to measure the performance, the *Precision* and *Recall* metrics were used:

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$
$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \quad (66)$$
$$\text{False Positive Rate} = 1 - \text{Precision}$$
$$\text{False Negative Rate} = 1 - \text{Recall}$$

As aforementioned, these metrics are evaluated for the proposed technique, as well as for other non-supervised techniques, to obtain a comparative framework and to validate the obtained results. The following techniques have been implemented and tested:

- *Region:* A region segmentation algorithm based on the *Otsu* method.
- *Edge*: A modern implementation of the *Canny* edge detector.
- *Edge-Region*: A combination of the two previous techniques. It is the one proposed in this paper.
- *SOM-Pix:* A *SOM* neural network based on the *RGB* intensity values of the image from the neighborhood of each pixel.
- *SOM-Avg:* A *SOM* neural network which input is the local average of the RGB values in a window centered in the neighborhood of the pixel.
- *SOM-Feat:* A *SOM* neural network that uses two different image features: the local average of the *RGB* values in a window centered in the neighborhood of the pixel and the standard deviation of the *RGB* values in the column and file of the pixel.

Three different versions of each technique were implemented; one without background information and another two where a background is used to normalize the image. In the first of these two versions, a static frame without fish was used to model the background. In the second one, the proposed dynamic background technique was used with every algorithm.

In every case, the images were enhanced with a standard contrast-limited adaptive histogram equalization technique.

When using the *SOM* techniques, a three-layer topology with 3 processing elements (neurons) in each layer was selected and a 3x3 window was used for each input. To mitigate the dependence of the results on the selected training patterns, all the networks were trained using a single image, and results were analysed using 3 different trainings. Additionally, the *SOM* network proposed in [38], based on different features from the current image and a static background, was considered in the comparative.

The results obtained after the representation and interpretation step for the different techniques are shown in Table 8. It can be observed that the results are strongly dependent on the background model. Only the proposed technique performed well without background modeling and, in general, dynamic background performed better than the static one.

It can also be noted that the worst results in terms of performance are obtained with the *Edge* technique. On the contrary, *SOM* models based on features obtain a good performance, but require an unacceptable execution time. Finally, it can be seen that although the *Region* technique obtained quite good results, the proposed technique achieved the same precision with the best recall, and without losing precision or increasing significantly the execution time.

Table 8: Performance of the selected techniques without tracking.

	Average Results	Precision	Recall	False Pos. Rate	False Neg. Rate	Time (s/frame)
No Backgnd.	SOM Pixel	0.43	0.29	0.57	0.71	1.38
	SOM Avg.	0.55	0.35	0.45	0.65	3.60
	SOM Feat.	0.63	0.26	0.37	0.74	11.63
	Edge	0.86	0.73	0.14	0.27	0.34
	Region	0.40	0.39	0.60	0.61	0.40
	Edge - Region	0.93	0.71	0.07	0.29	0.31
Static Backgnd.	SOM Pixel	0.91	0.49	0.09	0.51	1.55
	SOM Avg.	0.96	0.72	0.04	0.28	3.41
	SOM Feat.	0.96	0.62	0.04	0.38	11.22
	SOM [38]	0.85	0.69	0.15	0.31	3.37
	Edge	0.65	0.75	0.35	0.25	0.35
	Region	0.94	0.81	0.06	0.19	0.19
	Edge - Region	0.94	0.82	0.06	0.18	0.32
Dyn. Backgnd.	SOM Pixel	0.96	0.78	0.04	0.22	1.15
	SOM Avg.	0.96	0.61	0.04	0.39	11.62
	SOM Feat.	0.96	0.72	0.04	0.28	3.61
	Edge	0.67	0.78	0.33	0.22	0.32
	Region	0.95	0.78	0.05	0.22	0.17
	Edge - Region (proposed)	**0.95**	**0.82**	**0.05**	**0.18**	**0.31**

Once the representation and interpretation step is completed, fish detections are processed with the tracking algorithm. As explained above, this algorithm can operate as a filter, using the confirmed positions of the tracked fishes. The results obtained using this configuration with background modeling are shown in Table 9.

However, the tracking algorithm can also estimate the hidden positions of the fish. This is done by using a motion model and interpolation algorithms. Following this procedure, the results include both confirmed and estimated positions (Table 10).

Table 9: Performance with tracking operating as a filter

Selected techniques are used with background modeling after tracking and filtering steps and using only confirmed detections.

	Average Results	Precision	Recall	False Pos. Rate	False Neg. Rate	Time (sec./frame)
Static Backgnd.	SOM Pixel	0.97	0.81	0.03	0.19	1.14
	SOM Avg.	0.98	0.74	0.02	0.26	3.71
	SOM Feat.	0.97	0.64	0.03	0.36	11.97
	SOM [38]	0.88	0.75	0.12	0.25	3.27
	Edge	0.70	0.83	0.30	0.17	0.33
	Region	0.97	0.84	0.03	0.16	0.18
	Edge - Region	0.97	0.84	0.03	0.16	0.31
Dyn. Backgnd.	SOM Pixel	0.97	0.80	0.03	0.20	1.15
	SOM Avg.	0.98	0.74	0.02	0.26	3.67
	SOM Feat.	0.97	0.63	0.03	0.37	11.87
	Edge	0.75	0.85	0.25	0.15	0.33
	Region	0.97	0.81	0.03	0.19	0.17
	Edge - Region (proposed)	**0.98**	**0.85**	**0.02**	**0.15**	**0.31**

Table 10: Performance with filtering and interpolation tracking

Selected techniques are used with background modeling after tracking and filtering steps. Both confirmed and predicted detections are used.

	Average Results	Precision	Recall	False Pos. Rate	False Neg. Rate	Time (sec./frame)
8tatic Backgnd.	SOM Pixel	0.93	0.91	0.07	0.09	1.14
	SOM Avg.	0.95	0.86	0.05	0.14	3.71
	SOM Feat.	0.93	0.80	0.07	0.20	11.97
	SOM [38]	0.73	0.88	0.27	0.12	3.27
	Edge	0.44	0.93	0.56	0.07	0.33
	Region	0.94	0.94	0.06	0.06	0.18
	Edge - Region	0.94	0.94	0.06	0.06	0.31
Dyn. Backgnd.	SOM Pixel	0.94	0.91	0.06	0.09	1.15
	SOM Avg.	0.95	0.86	0.05	0.14	3.67
	SOM Feat.	0.94	0.80	0.06	0.20	11.87
	Edge	0.48	0.93	0.52	0.07	0.33
	Region	0.94	0.93	0.06	0.07	0.17
	Edge - Region (proposed)	**0.95**	**0.94**	**0.05**	**0.06**	**0.31**

The precision of the system is increased significantly if the algorithm is configured to operate only as a filter. However, the use of predicted positions improves the recall, without losing precision when compared with the previous step (Table 8).

It must be taken into account that some of the new false positives that appear when using predictions are not errors. In fact, they may reflect the position of the fish when it is not observable in the images and the fish center position has not been manually marked.

In conclusion, the proposed technique obtains the best performance. It achieves one of the lowest false positive rates and the lowest false negative one, with one of the best execution times. That confirms that proposed technique obtains very reliable results finding the true positions of the fish, with a high probability and detecting the fishes in most of the situations.

3.2 Tracking

As shown in the previous section, the proposed technique performs comparatively better than the other implemented methods in terms of precision and recall. However, the ability of the algorithm not only to detect fishes in images, but also to generate trajectories from fishes, still has to be tested. In this section, the ability of the system to observe fishes along time is studied, which implies measuring the accuracy in the task of assigning detections to fishes.

Although there is not a standard metric to perform this task, the effectively of tracking algorithms can be measured by analyzing and counting tracking errors. From a general point of view, this type of errors can be classified as follows:

- Type 1: The output trajectory of a detected fish contains isolated noise detections.
- Type 2: A fish is not detected and does not generate a trajectory.
- Type 3: A group of noise detections is classified as a new fish and a trajectory is generated for a non-existing fish.
- Type 4: Two or more trajectories are generated for a single fish. This can happen if a group of noise detections interfere with the trajectory of a tracked fish, or if a tracked fish is lost for a long period of time.
- Type 5: Two or more fishes interact during some time, causing occlusions and overlapping in the images. This results in the assignment of some of the fishes to the wrong trajectory.

Type 1 errors are reflected as false positives in the results of Table 8 to Table 10. Besides, Type 2 errors do not occur, in practice, since video sequences are long enough to ensure that every fish is detected.

On the other hand, Type 5 errors are impossible to detect by human operators, since several interacting fishes cannot be visually discriminated in the images. However, interaction among fishes in high-velocity areas was barely observed and it does not affect the analysis of fish behavior when occurring in resting zones.

Finally, longer sequences are required to analyze type 3 and 4 errors, and two new datasets have been created. The first dataset consists in a long sequence of 46,000 frames with a single fish moving in only one pool. The second one is formed by 10 sequences of 1,000 frames each, where 2 fishes interact in the same pool.

The obtained results can be seen in Table 11. They show a very low error rate, with a tracking error every 5000 frames or more. These results confirm that the

proposed system is suitable for obtaining fish trajectories from the recorded images.

Table 11: Summary of the Tracking test

Dataset	Type 3 errors	Type 4 errors	Error Rate (Per 1000 Frames)
Single fish	7	0	0.15
Two fishes	0	2	0.18

3.3 Biological Results

The proposed system was applied to 15 assays conducted for two years in the full-scale vertical slot fishway model located at the CEDEX laboratory (Figure 81). During the corresponding migration period, four different species (a total of 259 fishes) were tested: iberian barbel (*Luciobarbus bocagei*), mediterranean barbel (*Luciobarbus guiraonis*), iberian straight-mouth nase (*Pseudochondrostoma polylepis*) and brown trout (*Salmo trutta*), as shown in Figure 91. The recordings of each assay last approximately 12 hours, and the recording frequency is 25Hz.

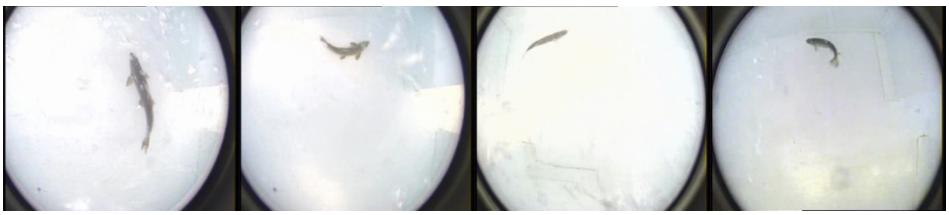

Figure 91: Fish species used in the assays.

From left to right: Iberian Barbel, Mediterranean Barbel, Iberian straight-mouth Nase and Brown Trout.

Overall, passage success during the experiments was low, regardless of species, and varied considerably with fish size (Table 12). In general, larger individuals presented a higher rate of success in ascending the entire fishway, relative to small specimens of the same species.

Table 12: Overall passage success during the experiments.

Specie	Size (cm)	Number of Fishes	Passage success (%)
Iberian Barbel	0 - 15	12	33.3%
	15 - 20	12	75.0%
	20 - 25	5	80.0%
	>25	34	41.2%
Mediterranean Barbel	0 - 15	6	0.0%
	15 - 20	8	0.0%
	20 - 25	11	54.6%
	>25	12	25.0%
Iberian straight-mouth Nase	0 - 15	61	9.8%
	20 - 25	34	41.2%
Brown Trout	0 - 15	5	0.0%
	15 - 20	43	14.0%
	20 - 25	14	42.9%
	25-30	2	100.0%
TOTAL		259	28.6%

On the other hand, the path chosen by fish moving from one pool to another and the specific resting zones actually exploited by the fish were identified. In the experiments, the individuals avoided high-velocity areas and used recirculation regions, in which velocity and turbulence levels are lower, to move within the pool and for resting before ascending through the higher velocity area of the slot. Thus, a preliminary analysis of the fish trajectories revealed that, when ascending the fishway, fishes spent more than 95% of the time in low velocity areas.

Besides, low-velocity areas were not frequented uniformly by fish, which stayed most frequently in the zone located just downstream from the slot and behind the small side baffle (zone A in Figure 92). The exploitation of low-velocity areas for the four species can be seen in Table 13. The frequency of use of resting zones is expressed as the proportion between the time spent in a specific one and the total resting time during the ascent.

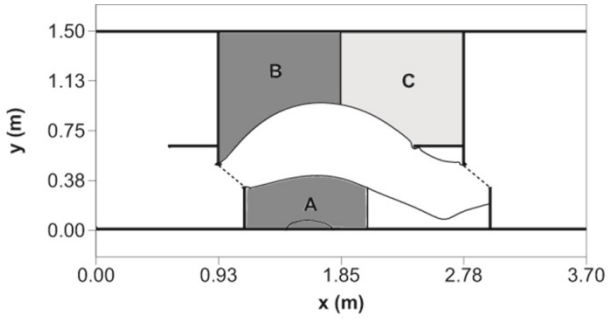

Figure 92: Location of the resting zones considered in this work.

Table 13: Exploitation of resting areas for the four species.
The resting zones are labeled as in Figure 92.

	Frequency of use (%)			Avg. resting time (s)
	A	B	C	
Iberian Barbel	68.5	28.5	2.9	161
Mediterranean Barbel	88.4	10.5	1.1	325
Iberian straight-mouth nase	99.8	0.0	0.2	269
Brown Trout	82.7	16.7	0.6	1271
TOTAL	84.1	15.1	0.8	585

Finally, the trajectory during the pool ascents was analyzed. In general, two modes of successful ascents were observed, depending on the location of the individual within the pool before traversing the slot and the area used to approach it (Figure 93). The results suggest that all the selected fish species tend to follow similar trajectories, and exploit the same flow regions during the ascent.

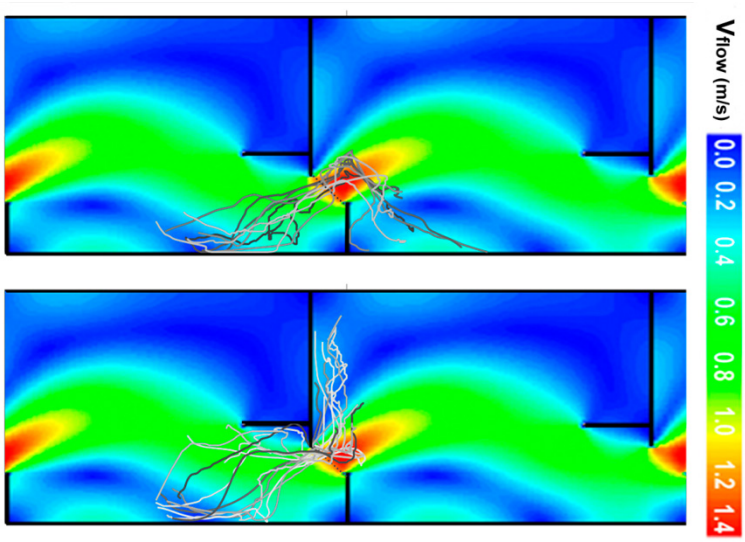

Figure 93: Examples of the two typical successful ascents used by fishes.

Besides, the observed speed, swimming speed and acceleration have been calculated as described in section 2.6. Figure 94 illustrates an example of results, in which velocities (in modulus) are represented as a function of the traveled distance. Additionally, average maximum speeds and acceleration can be seen in Table 14.

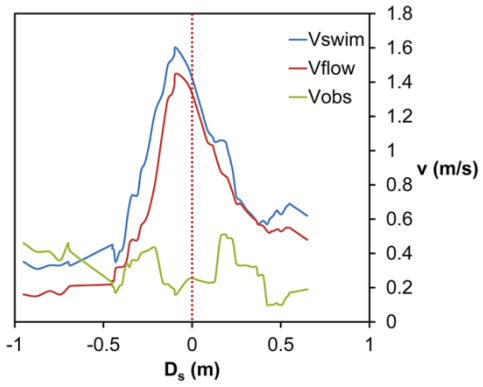

Figure 94: Example of velocities calculated for a fish

The shown velocities correspond to the fish which followed the trajectory shown in Figure 90. Ds is the traveled length from the slot section.

Table 14: Average maximum speeds and accelerations for the four species.

Specie	Swimming Speed (m/s)		Acceleration (m/s2)	
	Avg. Maximum	Std. Deviation	Avg. Maximum	Std. Deviation
Iberian Barbel	1.51	0.27	1.13	0.60
Mediterranean Barbel	1.51	0.25	0.95	0.60
Iberian straight-mouth nase	1.52	0.26	1.08	0.54
Brown Trout	1.60	0.25	1.31	0.74

4 Conclusions

In this thesis, a data acquisition system with video monitoring to record experiments on a full-scale fishway model has been designed and built. This system allows the observation and monitoring of fish in an accurate and effective way.

In addition, a new technique has been developed to automatically analyze fish behavior in fishways. It uses computer vision techniques and algorithms to detect and track fishes in video sequences recorded by the camera system integrated in the fishway.

More specifically, it employs a combination of background modeling, edge and region analysis to detect fishes, taking advantage of the Kalman filter to obtain the trajectory of one or multiple individuals.

The proposed technique has been extensively tested and compared with different standard methods, obtaining the best balance of precision, recall and execution time.

The system has been applied to 15 assays performed along two years and using more than 250 living fishes. It has provided valuable information regarding fish

behavior, including fishway efficiency, swimming trajectories, swimming velocities and accelerations, resting times and preferential resting areas.

Some of the finds of this applications were published in different media, the main of these contributions are listed below:

A. Rodríguez, J. Rabuñal, J. Dorado, and J. Puertas, "Overflow Channel Study using Image Procesing," presented at the International conference on Image Processing & Communications (IPC), Bydgoszcz, Poland, 2009.

A. Rodriguez, M. Bermudez, J. Rabuñal, J. Puertas, J. Dorado, and L. Balairon, "Optical Fish Trajectory Measurement in Fishways through Computer Vision and Artificial Neural Networks," Journal of Computing in Civil Engineering, vol. 25, pp. 291-301, 2011.

J. Puertas, L. Cea, M. Bermudez, L. Pena, A. Rodriguez, J. Rabuñal, et al., "Computer application for the analysis and design of vertical slot fishways in accordance with the requirements of the target species," Ecological Engineering, vol. 48, pp. 51-60, 2012.

A. Rodriguez, J. Rabuñal, M. Bermudez, and A. Pazos, "Automatic Fish Segmentation on Vertical Slot Fishways Using SOM Neural Networks," presented at the International work Conference on Artificial Neural Networks (IWANN), Tenerife, España, 2013.

A. Rodriguez, M. Bermúdez, J. Rabuñal, and J. Puertas, "Fish tracking in vertical slot fishways using computer vision techniques," 2014. (Being Processed)

Although further research is needed, the results obtained with this application, together with the results of upcoming assays, may contribute to develop robust

guidelines for future fishway designs and to establish more realistic criteria for the evaluation of biological performance of the current designs.

Using no way as a way,
using no limitations as a limitation
Lee Jun-Fan

VII. Conclusions and Future Work

This PhD thesis is focused on the application of computer vision (*CV*) techniques in the area of Civil Engineering, assuming that these techniques can replace some of the traditional instrumentation used in Civil Engineering experiments, obtaining more flexible, cheaper, and accurate results.

To validate this hypothesis, the main objective of this thesis was to develop a methodology to develop Computer Vision systems in Civil Engineering.

This goal has been achieved by studying the characteristics of Computer Vision and Civil Engineering and by identifying the tasks, relationships and components required in a general system and then proposing a development methodology.

Therefore, the cycle of general purpose *CV* system development using systematic and efficient processes have been carried out, explaining the most important and specific tasks and techniques required for its application in Civil Engineering.

According to this methodology, two *CV* systems were developed, being significant contributions to the Civil Engineering field.

The first of the developed systems consists in a new technique to analyse the strains in material strength tests using recorded images of the experiment.

The proposed technique is based on obtaining the correspondence of the image regions called blocks. It has advantages of robustness and flexibility over other *CV* algorithms: It allows imposing different deformation models in various stages of the algorithm and using different similarity metrics, subpixel algorithms or post-processing filters, allowing to impose different degrees of freedom in diverse stages of the algorithm.

The proposed technique has been compared with different *CV* techniques and with traditional instrumentation to measure strength tests in different assays. It obtained the best accuracy when compared with other *CV* techniques and with traditional instrumentation. It obtained similar results in accuracy, but having several advantages: the cost being lower and the measurement more flexible.

The second developed system was an entire computer vision system for analysing fishes inside vertical slot fishways with different computer vision techniques.

The proposed system is composed by a physical data acquisition system, designed to be integrated in a fishway and to record experiments, and a new technique to analyse automatically the fish behavior from the images recorded with the acquisition system.

The proposed algorithm uses a combination of background modeling, edge and region analysis to detect fishes, taking advantage of the Kalman filter to obtain the trajectory of one or multiple individuals.

The proposed technique has been extensively tested and compared with different standard methods, obtaining the best results in terms of precision, recall and execution time.

The system has been applied to 15 assays performed along two years and using more than 250 living fishes. It has provided valuable information regarding fish behavior, including fishway efficiency, swimming trajectories, swimming velocities and accelerations, resting times and preferential resting areas.

With these and the upcoming results, it is expected that the developed system will contribute to develop robust guidelines for future fishway designs and to establish more realistic criteria for the evaluation of the current ones.

The developed systems show that the proposed methodology is useful and validates the hypotheses of this thesis.

It is expected that the presented work may be extended in the future, establishing more precise guidelines to design and construct *CV* systems and offering a practical guide for civil engineers.

In future stages, the proposed methodology may be applied in more scenarios and extended to cover more requirements of civil engineers and to include 3D scenarios and different optical devices.

To see things in the seed is genius

Lao Tzu

VIII. Bibliography

[1] T. Brosnan and D. W. Sun, "Improving quality inspection of food products by computer vision-a review," *Journal of Food Engineering,* vol. 61, pp. 3-16, 2004.

[2] D. Marr, *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco: W. H. Freeman 1982.

[3] J. M. Jolion, "Computer Vision Methodologies," *CVGIP: Image Understanding,* vol. 59, pp. 53-71, 1994.

[4] F. H. Shih, *Image Processing and Pattern Recognition*. United States: Wiley 2010.

[5] R. Fisher, S. Perkins, A. Walker, and E. Wolfart. (2004, 2012-03-21). *Image processing Learning resources*. Available: http://homepages.inf.ed.ac.uk/rbf/HIPR2/hipr_top.htm

[6] X. Qian, J. Wang, S. Guo, and Q. Li, "An active contour model for medical image segmentation with application to brain CT image," *Medical Physics,* vol. 40, pp. -, 2013.

[7] H. Moftah, A. Azar, E. Al-Shammari, N. Ghali, A. Hassanien, and M. Shoman, "Adaptive k-means clustering algorithm for MR breast image segmentation," *Neural Computing and Applications,* pp. 1-12, 2013/06/19 2013.

[8] J. Romanowski, T. Nowak, P. Najgebauer, and S. Litwiński, "Improved X-ray Edge Detection Based on Background Extraction Algorithm," in *Artificial Intelligence and Soft Computing*. vol. 7895, L. Rutkowski, M. Korytkowski, R. Scherer, R. Tadeusiewicz, L. Zadeh, and J. Zurada, Eds., ed: Springer Berlin Heidelberg, 2013, pp. 309-319.

[9] A. Schenk, G. Prause, and H.-O. Peitgen, "Efficient Semiautomatic Segmentation of 3D Objects in Medical Images," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2000*. vol. 1935, S. Delp, A. DiGoia, and B. Jaramaz, Eds., ed: Springer Berlin Heidelberg, 2000, pp. 186-195.

[10] W. K. Pratt, *Digital Image Processing*, 4th ed.: Wiley, 2007.

[11] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 2nd ed.: Prentice Hall, 2002.

[12] A. Rodriguez, "A mirada artificial, ¿Pode un computador ver?," *Boletín das Ciencias (ENCIGA),* vol. 72, pp. 19-25, 2011.

[13] R. Szeliski, *Computer Vision: Algorithms and Applications*: Springer, 2011.

[14] A. H. Pierce, "The illusions of the kindergarten patterns," *Psych. Review,* vol. 5, pp. 233-253, 1898.

[15] E. H. Adelson. (2005, 2014/01). *Checkershadow Illusion*. Available: http://web.mit.edu/persci/people/adelson/checkershadow_illusion.html

[16] B. Lingelbach and W. H. Ehrenstein, "Das Hermann-Gitter und die Folgen," *DOZ,* vol. 5, pp. 13-20, 2002.

[17] R. N. Shepard, "Elephant illusion," in *Mind sights: Original visual illusions, ambiguities, and other anomalies*, ed New York: H. Freeman 1990.

[18] R. Szeliski, "Computer Vision Timeline," in *Computer Vision: Algorithms and Applications*, ed: Springer, 2011.

[19] L. G. Shapiro and G. C. Stockman, *Computer Vision*: Prentice Hall, 2001.

[20] T. Morris, *Computer Vision and Image Processing*: Palgrave Macmillan, 2004.

[21] B. Jähne and H. Haußecker, *Computer Vision and Applications, a Guide for Students and Practitioners*: Academic Press, 2000.

[22] A. Rodriguez, C. Fernández-Lozano, J. Dorado, and J. Rabuñal, "2-D Gel Electrophoresis Image registration using Block-Matching Techniques and Deformation Models," *Analytical Biochemistry,* 2014.

[23] C. Steger, M. Ulrich, and C. Wiedemann, *Machine Vision Algorithms and Applications*: Wiley-VCH, 2008.

[24] M. Graves and B. Batchelor, *Machine Vision for the Inspection of Natural Products*: Springer, 2003.

[25] C. G. Relf, *Image Acquisition and Processing with LabVIEW*: CRC Press, 2004.

[26] R. I. Hartley and A. Zisserman, *Multiple View Geometry*. Cambridge: Cambridge University Press, 2004.

[27] Z. Zhang, "Flexible Camera Calibration By Viewing a Plane From Unknown Orientations," presented at the International Conference on Computer Vision (ICCV), Kerkyra, Greece, 1999.

[28] A. Rodríguez , C. Fernández-Lozano, and J. Rabuñal, *Non-rigid motion estimation using Block-Matching Techniques*: LAP Lambert Academic Publishing, 2012.

[29] A. Rodríguez , C. Fernández-Lozano, and J. Rabuñal, "Spot matching in 2D electrophoresis experiments," *International journal of Data Analysis and Strategies,* vol. 5, pp. 198-213, 2013.

[30] L. M. Lye, "Design of Experiments in Civil Engineering: are we still in 1920's," presented at the Annual Conference of the Canadian Society for Civil Engineering, Montréal, Québec, Canada, 2002.

[31] A. Rodriguez, J. Rabuñal, M. Bermudez, and A. Pazos, "Detection of Fishes in Turbulent Waters Based on Image Analysis," presented at the International Work-Conference on the Interplay between Neural and Artificial Computation (IWINAC), Palma de Mallorca, España, 2013.

[32] A. Rodriguez, C. Fernandez-Lozano, J.-A. Seoane, J. Rabuñal, and J. Dorado, "Analysis of Deformation Processes Using Block-Matching Techniques," presented at the International Conference on Computer Vision Theory and Applications, Roma, Italia, 2012.

[33] A. Rodriguez, J. Rabuñal, J. L. Perez, and F. Martinez-Abella, "Study of Strenght Tests with Computer Vision Techniques," presented at the International Work-Conference on the interplay between Natural and Artifial Computation (IWINAC), Islas Canarias, 2011.

[34] C. A. Glasbey and K. V. Mardia, "A review of image warping methods " *Journal of applied statistics,* vol. 25, pp. 155-171, 1998.

[35] P. R. V., T. R. K., D. Y.B., Y. K.M., M. C., and S. M.V.L., "Performance Evaluation of Object Tracking Technique Based on Position Vectors " *International Journal of Image Processing (IJIP),* vol. 7, pp. 124-131, 2013.

[36] M. K. Jalloul and M. A. Al-Alaoui, "A novel parallel motion estimation algorithm based on Particle Swarm Optimization," in *Signals, Circuits and Systems (ISSCS), 2013 International Symposium on*, 2013, pp. 1-4.

[37] D. A. Fosyth and J. Ponce, *Computer Vision: A Modern Approach*: Prentice Hall, 2003.

[38] A. Rodriguez, M. Bermudez, J. Rabuñal, J. Puertas, J. Dorado, and L. Balairon, "Optical Fish Trajectory Measurement in Fishways through Computer Vision and Artificial Neural Networks," *Journal of Computing in Civil Engineering,* vol. 25, pp. 291-301, 2011.

[39] A. Cunningham. (2011, 2014/01). *Choosing Camera Lens Filters, The official blog of Peter Lik*. Available: http://www.peterlikexposed.com

[40] A. Freire, J. A. Seoane, A. Rodriguez, C. Ruiz Romero, G. López-Camos, and J. Dorado, "A Block Matching based technique for the analysis of 2D gel images," *Studies in health technology and informatics,* vol. 160, pp. 1282-1286, 2010.

[41] A. Rodriguez, M. Bermudez, D. Rivero, M. Gestal, and J. Puertas, "Seguimiento Visual de Peces en Escalas de Hendidura Vertical," presented at the Congreso nacional sobre Metaheurísticas, Algoritmos Evolutivos y Bioinspirados (MAEB), Albacete, España, 2012.

[42] K. Singh and S. R. Ahamed, "Systolic array based architecture for DS fast motion estimation algorithm," in *Image Information Processing (ICIIP), 2013 IEEE Second International Conference on*, 2013, pp. 335-339.

[43] S. Robertson, E. Weiss, and G. Hugo, "TU-C-141-03: Block-Matching Registration for Localization of Locally-Advanced Lung Tumors During Image-Guided Radiotherapy " *Medical Physics,* vol. 40, pp. 434-434, 2013.

[44] E. Monteiro, M. Maule, F. Sampaio, C. Diniz, B. Zatt, and S. Bampi, "Real-time block matching motion estimation onto GPGPU," in *Image Processing (ICIP), 2012 19th IEEE International Conference on*, 2012, pp. 1693-1696.

[45] E. Cuevas, D. Zaldívar, M. Pérez-Cisneros, and D. Oliva, "Block-matching algorithm based on differential evolution for motion estimation," *Engineering Applications of Artificial Intelligence,* vol. 26, pp. 488-498, 1// 2013.

[46] F. L. Bookstein, "Principal warps: thin-plate splines and the decomposition of deformations," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI),* vol. 11, pp. 567-585, 1989.

[47] J. Puertas, L. Cea, M. Bermudez, A. Rodriguez, and J. Rabuñal, "Computer application for the analysis and design of vertical slot fishways in acordance with the requirements of the target species," presented at the International congress on the biology of fish, Barcelona, Spain, 2010.

[48] E. Cuevas, D. Zaldívar, M. Pérez-Cisneros, H. Sossa, and V. Osuna, "Block matching algorithm for motion estimation based on Artificial Bee Colony (ABC)," *Applied Soft Computing,* vol. 13, pp. 3047-3059, 6// 2013.

[49] A. Rodriguez, J. Rabuñal, J. L. Perez, and F. Martinez-Abella, "Optical Analysis of Strength Tests Based on Block-Matching Techniques," *Computer-Aided Civil and Infrastructure Engineering,* vol. 27, pp. 573–593, 2012.

[50] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," presented at the Imaging Understanding Workshop, 1981.

[51] B. K. P. Horn and S. B.G., "Determining optical flow," *Artificial Intelligence,* vol. 17, pp. 185–203, 1981.

[52] D. J. Heeger, "Optical flow using spatiotemporal filters," *International Journal of Computer Vision (IJCV),* vol. 1, pp. 279–302, 1988.

[53] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking " *IEEE Transactions on Signal Processing,* vol. 50, pp. 174-188 2002.

[54] A. Rodriguez, M. Bermudez, J. Rabuñal, and J. Puertas, "Optical fish tracking in fishways using Neural Networks," presented at the International Conference on Image, Signal and Vision Computing (ICISVC), Paris, France, 2010.

[55] DoITPoMS. (2014/01). *Introduction to Mecanical testing, Univertsity of Cambridge*. Available: http://www.doitpoms.ac.uk/tlplib/mechanical-testing

[56] Oceanlab. (2014/01). *Oceanlab, University of Aberdeen*. Available: http://www.oceanlab.abdn.ac.uk/

[57] R. Raskar and J. Tumblin, *Computational Photography: Mastering New Techniques for Lenses, Lighting, and Sensors*. Massachusetts: A K Peters, 2010.

[58] M. J. Hannah, "Computer Matching of Areas in Stereo Images," Stanford University, 1974.

[59] R. C. Bolles, H. H. Baker, and M. J. Hannah, "The JISCT stereo evaluation," presented at the Image Understanding Workshop, 1993.

[60] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *International Journal of Computer Vision (IJCV),* vol. 47, pp. 7-42, 2002.

[61] J. A. Worthey, "Shape Recovery. ," in *Physics-Based Vision: Principles and Practice.* vol. 18, L. B. Wolff, S. A. Shafer, G. E. Healey, and A. K. Peters, Eds., ed: Wellesley, MA, 1993, pp. 221-223.

[62] R. Szeliski and H. Y. Shum, "Creating full view panoramic image mosaics and texture-mapped models," presented at the ACM SIGGRAPH, 1997.

[63] K. Lillywhite, D.-J. Lee, B. Tippetts, and J. Archibald, "A feature construction method for general object recognition," *Pattern Recognition,* vol. 46, pp. 3300-3314, 12// 2013.

[64] E. P. Simoncelli and E. H. Adelson, "Computing Optical Flow Distributions using Spatio-temporal filters," MIT Media Laboratory Vision and Modeling1990.

[65] H. Y. Shum, S. C. Chan, and S. B. Kang, *Image-Based Rendering*. New York, NY: Springer, 2007.

[66] J. Ponce, T. Berg, M. Everingham, D. A. Forsyth, M. Hebert, S. Lazebnik, *et al.*, "Dataset issues in object recognition," in *Toward Category-Level Object Recognition*, J. Ponce, M. Hebert, C. Schmid, and A. Zisserman, Eds., ed New York: Springer 2006, pp. 29–48.

[67] I. o. C. Engineering. (2014/01). *Institution of Civil Engineering*. Available: http://www.ice.org.uk

[68] A. Rodríguez, J. Rabuñal, J. Dorado, and J. Puertas, "Overflow Channel Study using Image Procesing," presented at the International conference on Image Processing & Communications (IPC), Bydgoszcz, Poland, 2009.

[69] M. Shinoda and R. J. Bathurst, "Strain measurement of geogrids using a video-extensometer technique," *Geotechnical testing journal,* vol. 27, pp. 456-463, 2004.

[70] E. Fauster, P. Schalk, and P. L. O'Leary, "Evaluation and calibration methods for the application of a video-extensometer to tensile testing of polymer materials," 2005, pp. 187-198.

[71] G. Iglesias, I. O., A. Castro, J. R. Rabuñal, and J. Dorado, "Computer vision applied to wave flume measurements," *Ocean Engineering,* vol. 35, pp. 1073-1079, 2009.

[72] Bosh. (2014/01). *Design world: Bosch Rexroth helps outfit new automation lab at Illinois state university*. Available: http://www.designworldonline.com/bosch-rexroth-helps-outfit-new-automation-lab-at-illinois-state-university

[73] C. P. Lu, G. D. Hager, and E. Mjolsness, "Fast and Globally Convergent Pose Estimation from Video Images," *IEEE Transactions on Pattern and Machine Intelligence,* vol. 22, pp. 610-622, 2000.

[74] J. Weng, P. Cohen, and M. Herniou, "Camera calibration with distortion models and accuracy evaluation.," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI),* vol. 14, pp. 965-980, 1992.

[75] OPENCV. (2013/12). *Open Source Computer Vision*. Available: http://opencv.org/

[76] V. Bockaer. (1998, 2013/12). *Dpreview, Pixel Quality*. Available: http://www.dpreview.com/glossary/camera-system/pixel-quality

[77] B. Singh, R. S. Mishra, and P. Gour, "Analysis of Contrast Enhancement Techniques for Underwater Image," *International Journal of Computer Technology and Electronics Engineering,* vol. 1, pp. 190-194, 2011.

[78] A. Rodriguez and J. Rabuñal, "Measuring Deformations in Biomedicine with Block-Matching Techniques: A Survey," 2014.

[79] Infaimon. (2011, 2013/12). *Vision Artificial Infaimon SL, Catálogo de Visión Artificial*. Available: http://www.infaimon.com/catalog/catalog_files.php?productid=16896&fileid=1377&sl=AO

[80] EM. (2014/01). *Electro-Magnetic Spectrum and light*. Available: http://9-4fordham.wikispaces.com/Electro+Magnetic+Spectrum+and+light

[81] NI. (2008, 2013/12). *National Instruments, A practical guide to machine vision lightning*. Available: http://zone.ni.com/devzone/cda/tut/p/id/6901

[82] Microscan. (2012, 2014/01). *Microscan Systems Inc, Machine Vision Lighting Training*. Available: http://www.microscan.com/trainingandresources/lighting.aspx

[83] B. Wieneke, "Method for determining flow conditions," ed: Google Patents, 2013.

[84] Aven. (2012, 2014/01). *Aven Inc, Machine Vision Illumination Guide.* . Available: http://www.aventools.com/pdf/mv_illumination_guide.pdf

[85] ThinFilm. (2014/01). *Optical Filters, Newport Thin Film Laboratory*. Available: http://newportlab.com/services/filters/

[86] B. Wieneke, "Method of determining a three-dimensional velocity field in a volume," ed: Google Patents, 2008.

[87] Y. Song, D. Treanor, A. J. Bulpitt, N. Wijayathunga, N. Roberts, R. Wilcox, *et al.*, "Unsupervised Content Classification Based Nonrigid Registration of Differently Stained Histology Images," *Biomedical Engineering, IEEE Transactions on,* vol. 61, pp. 96-108, 2014.

[88] O. Commowick, N. Wiest-Daessle, and S. Prima, "Block-matching strategies for rigid registration of multimodal medical images," in *Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on*, 2012, pp. 700-703.

[89] S. R. Cruz-Ramírez, Y. Mao, T. Arao, T. Tacuba, and K. Ohana, "Vision-based Hierarchical Recognition for Dismantling Robot Applied to Interior Renewal of Buildings," *Computer-Aided Civil and Infrastructure Engineering,* vol. 26, pp. 336-355, 2011.

[90] A. Abbas, M. E. Kutay, H. Azari, and R. Rasmussen, "Three-Dimensional Surface Texture Characterization of Portland Cement Concrete Pavements," *Computer-Aided Civil and Infrastructure Engineering,* vol. 22, pp. 197-209, 2007.

[91] L. Ying and E. Salari, "Beamlet Transform Based Technique for Pavement Image Processing and Classification," *Computer-Aided Civil and Infrastructure Engineering,* vol. 25, pp. 572-580, 2010.

[92] A. Freire, J.-A. Seoane, V. Aguiar, A. Rodriguez, and J. Rabuñal, "Protein identification over 2D gel images using computer vision techniques," presented at the ll Xornada galega de Bioinformática, Santiago de Compostela, España, 2009.

[93] C. Je and H.-M. Park, "Optimized hierarchical block matching for fast and accurate image registration," *Signal Processing: Image Communication,* vol. 28, pp. 779-791, 8// 2013.

[94] INS. (2011, 2013/12). *Kinston*. Available: http://www.instron.com

[95] G. Zahnd, M. Orkisz, A. Sérusclat, P. Moulin, and D. Vray, "Evaluation of a Kalman-based block matching method to assess the bi-dimensional motion of the carotid artery wall in B-mode ultrasound sequences," *Medical Image Analysis,* vol. 17, pp. 573-585, 2013.

[96] P. J., V. Lepetit, and P. Fu, "Fast Non-Rigid Surface Detection, Registration and Realistic Augmentation," *International Journal of Computer Vision (IJCV),* vol. 76, pp. 109-122, 2007.

[97] K. Chivers and W. Clocksin, "Inspection of Surface Strain in Materials Using Optical Flow," presented at the British Machine Vision Conference, 2000.

[98] M. Raffel, C. Willert, and J. Kompenhans, *Particle Image Velocimetry, a Practical Guide*: Springer, 2000.

[99] M. Raffel, C. Willert, and J. Kompenhans, *Particle Image Velocimetry, a Practical Guide*, 2nd ed.: Springer, 2007.

[100] A. Fusiello and V. Roberto, "Symmetric stereo with multiple windowing," *International Journal of Pattern Recognition and Artificial Intelligence,* vol. 14, pp. 1053–1066, 2000.

[101] D. Schwarz and T. Kasparek, "Multilevel Block Matching technique with the use of Generalized Partial Interpolation for Nonlinear Intersubject Registration of MRI Brain Images," *European Journal for Biomedical Informatics,* vol. 1, pp. 90-97, 2006.

[102] T. Amiaz, E. Lubetzky, and M. Kiryati, "Coarse to over-fine optical flow estimation," *Pattern Recognition,* vol. 40, pp. 2496-2503, 2007.

[103] J. Liangbao, C. Jiao, C. Xuehong, and C. Rui, "Hierarchical Support-weight Block Matching Approach in Depth Extraction," *International Journal of Digital Content Technology and its Applications,* vol. 5, pp. 30-38, 2011.

[104] Y. Luo and M. Celenk, "A hybrid block-matching approach to motion estimation with adaptive search area," presented at the 15th International Conference on Systems, Signals and Image Processing (IWSSIP), 2008.

[105] H.-S. Oh and H.-K. Lee, "Block matching algorithm based on an adaptative reduction of the search area for motion estimation," *Real-Time Imaging,* vol. 6, pp. 407-414, 2000.

[106] D. Aboutajdine and F. Essannouni, "Fast block matching algorithms using frequency domain. International Conference on Multimedia Computing and Systems," presented at the International Conference on Multimedia Computing and Systems (ICMCS), 2011.

[107] F. F. J. Schrijer and F. Scarano, "On the Stabilization and Spatial Resolution of Iterative PIV Interrogation," presented at the 13th Symposium on Applications of Laser Techniques to Fluid Mechanics, 2006.

[108] A. Rodriguez, C. Fernandez-Lozano, J.-A. Seoane, J. Rabuñal, and J. Dorado, "Motion Estimation in Real Deformation Processes Based on Block-Matching Techniques," presented at the IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), 2011.

[109] U. Malsch, C. Thieke, P. E. Huber, and R. Bendl, "An enhanced block matching algorithm for fast elastic registration in adaptive radiotherapy," *Physics in Medicine and Biology,* vol. 51, pp. 4789-4806, 2006.

[110] I. Patras, E. A. Hendriks, and R. L. Lagendijk, "Probabilistic Confidence Measures for Block Matching Motion Estimation," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 17, pp. 988-995, 2007.

[111] S. Roth and M. Black, "On the Spatial Statistics of Optical Flow," *International Journal of Computer Vision (IJCV),* vol. 74, pp. 33-50, 2007.

[112] X. Yuan and X. Shen, "Block Matching Algorithm Based on Particle Swarm Optimization for Motion Estimation," presented at the International Conference on Embedded Software and Systems (ICESS), 2008.

[113] D. Zhāng and G. Lu, "An edge and color oriented optical flow estimation using block matching," presented at the International Conference on Signal Processing (ICSP), 2000.

[114] X. Shi and J. Chen, "Image mosaics algorithm based on feature-block matching," presented at the International Conference on Multimedia Technology (ICMT), 2011.

[115] M. A. Sekeh, M. A. Maarof, M. F. Rohani, and M. Motiei, "Efficient image block matching algorithm with two layer feature extraction," in *7th International Conference on Information Technology in Asia (CITA)*, 2011, pp. 1-5.

[116] I. Stuke, T. Aach, E. Barth, and C. Mota, "Multiple-Motion-Estimation by Block-matching using MRF," *International Journal of Computer & Information Science (IJCIS),* vol. 5, 2004.

[117] B. Xiong and C. Zhu, "A New Multiplication-Free Block Matching Criterion," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 18, pp. 1441-1446, 2008.

[118] J. Lu and M. L. Liou, "A simple and efficient search algorithm for block-matching motion estimation," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 7, pp. 429-433, 1997.

[119] S. Kant, P. S. S. B. K. Gupta, R. Korana, and M. Liar, "Efficient search algorithms for block-matching motion estimation," presented at the IEEE International Conference on Multimedia and Expo (ICME), 2008.

[120] K. Liu, Q. Qiu, and Z. Zhang, "A novel fast motion estimation algorithm based on block-matching," presented at the Cross Strait Quad-Regional Radio Science and Wireless Technology Conference (CSQRWC), 2011.

[121] X. Ren, "Local Grouping for Optical Flow," presented at the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 2008.

[122] C. Rhemann, A. Hosni, C. Rother, and M. Gelautz, "Fast cost-volume filtering for visual correspondence and beyond," presented at the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 2011.

[123] S. X. Ju, M. J. Black, and A. D. Jepson, "Skin and bones: Multi-layer, locally affine, optical flow and regularization with transparency," presented at the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1996.

[124] M. H. Chan, Y. B. Yu, and A. G. Constantinides, "Variable size block matching motion compensation with applications to video coding," in *IEE Proceedings I Communications, Speech and Vision*, 1990, pp. 205-212.

[125] T. Kanade and M. Okutomi, "A stereo matching algorithm with an adaptive window: Theory and experiment," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI),* vol. 16, pp. 920-932, 1994.

[126] G. R. Martin, R. A. Packwood, and I. Rhee, "Variable size block matching motion estimation with minimal error," presented at the SPIE Digital Video Compression: Algorithms and Technologies, 1996.

[127] S. B. Kang, R. Szeliski, and J. Chai, "Handling occlusions in dense multi-view stereo," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* vol. 1, pp. 103–110, 2001.

[128] O. Veksler, "Fast variable window for stereo correspondence using integral images," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* vol. 1, pp. 556–561, 2003.

[129] M. Gong, R. Yang, L. Wang, and M. Gong, "A performance study on different cost aggregation approaches used in real-time stereo matching," *International Journal of Computer Vision (IJCV),* vol. 75, pp. 283–296, 2007.

[130] K.-J. Yoon and S. Kweon, "Adaptive support-weight approach for correspondence search," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI),* vol. 24, pp. 650–656, 2006.

[131] J. Delon and B. Rougé, "Small baseline stereovision," *Journal of Mathematical Imaging and Vision,* vol. 28, pp. 209–223, 2007.

[132] C. Schmid and A. Zisserman, "The geometry and matching of lines and curves over multiple views," *Internationa Journal of Computer Vision (IJCV),* vol. 40, pp. 199–234, 2000.

[133] M. Awanish, K. V. N. Tiwari, and T. Amod, "Design and Development of Scaled Value Criterion for Video Compression using Block Matching

Techniques," *International Journal of Computer Applications,* vol. 64, pp. 1-8, 2013.

[134] N. Wiest-Daessle and S. Prima, "Block-matching strategies for rigid registration of multimodal medical images," presented at the IEEE International Symposium on Biomedical Imaging (ISBI), 2012.

[135] B. Karaçali, "Information Theoretic Deformable Registration Using Local Image Information," *International Journal of Computer Vision (IJCV),* vol. 72, pp. 219-237, 2007.

[136] G. Dong and M. Xie, "Color clustering and learning for image segmentation based on neural networks," *IEEE transactions on neural networks,* vol. 16, pp. 925-936, 2005.

[137] M. Arevalillo-Herraez, F. J. Ferri, and J. Domingo, "Learning Combined Similarity Measures from User Data for Image Retrieval," presented at the International Conference on Pattern Recognition (ICPR), 2008.

[138] J. Wachs, H. Stern, T. Burks, and V. Alchanatis, "Multi-modal Registration Using a Combined Similarity Measure," *Advances in Soft Computing,* vol. 52, pp. 159-168, 2009.

[139] J. Morales, F. J. Rodrigo, R. Verdú, and J. L. Sancho, "A combined similarity measure for image registration and fusion," *IADAT Journal of Advanced Technology on Imaging and Graphics,* vol. 1, pp. 36-38, 2005.

[140] GVRL. (2012/09). *Graphics and Vision Research Laboratory, Department of Computer Science, University of Otago*. Available: http://www.cs.otago.ac.nz

[141] J. P. Pluim, J. B. Maintz, and M. A. Viergever, "Mutual information based registration of medical images: a survey," *IEEE Transactions on Medical Imaging,* vol. 22, pp. 986-1004, 2003.

[142] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal,* vol. 27, pp. 379–423, 623–656, 1948.

[143] E. Geoffrey and D. Kostas, "Image Registration Using Mutual Information," University of Pennsylvania Department of Computer and Information Science2000.

[144] M. Shimizu and M. Okutomi, "Sub-Pixel Estimation Error Cancellation on Area-Based Matching," *International Journal of Computer Vision (IJCV),* vol. 63, pp. 207-224, 2005.

[145] H. B. Nielsen, "Damping Parameter in Marquardt's Method," University of Denmark IMM-REP-1999-05, 1999.

[146] N. Arad, N. Dyn, D. Reisfeld, and Y. Yeshurun, "Image Warping by Radial Basis Functions: Application to Facial Expressions," *CVGIP: Graphical Models and Image Processing,* vol. 1, pp. 161-172, 1994.

[147] A. Rodriguez, J. Rabuñal, M. Bermudez, and A. Pazos, "Automatic Fish Segmentation on Vertical Slot Fishways Using SOM Neural Networks," presented at the International work Conference on Artificial Neural Networks (IWANN), Tenerife, España, 2013.

[148] MVCP. (2013/12). *Middlebury computer vision pages*. Available: http://vision.middlebury.edu

[149] B. McCane, K. Novins, D. Crannitch, and B. Galvin, "On Benchmarking Optical Flow," *Computer Vision and Image Understanding,* vol. 84, pp. 126-143, 2001.

[150] D. J. Heeger, "Model for the extraction of image flow," *Journal of the Optical Society of America A: Optics, Image Science, and Vision,* vol. 4, pp. 1455–1471, 1987.

[151] I. Austvoll, "A Study of the Yosemite Sequence Used as a Test Sequence for Estimation of Optical Flow," *Lecture Notes in Computer Science,* vol. 3540, pp. 659-668, 2005.

[152] S. Baker, D. Scharstein, and J. P. Lewis, "A database and evaluation methodology for optical flow," *International Conference on Computer Vision (ICCV),* vol. 92, pp. 1-31, 2007.

[153] D. Sun, S. Roth, J. P. Lewis, and M. J. Black, "Learning Optical Flow," presented at the European Conference on Computer Vision (ECCV), 2008.

[154] D. Sun, S. Roth, and M. J. Black, "Secrets of Optical flow estimation and their principles," presented at the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 2010.

[155] PIV. (2013/12). *Particle image Velocimetry*. Available: http://www.piv.de

[156] Z. Deng, C. M. Richmond, G. R. Guest, and R. P. Mueller, "Study of Fish Response Using Particle Image Velocimetry and High-Speed, High-Resolution Imaging," US Department of Energy, Technical Report2004.

[157] D. Sun. (2013/12). *Dewing Sun Research page, Computer Science Department, Brown University*. Available: http://www.cs.brown.edu/~dqsun

[158] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert, "High Accuracy Optical Flow Estimation Based on a Theory for Warping," presented at the European Conference on Computer Vision (ECCV), 2004.

[159] N. Papenberg, A. Bruhn, T. Brox, S. Didas, and J. Weickert, "Highly Accurate Optic Flow Computation with Theoretically Justified Warping," *International Journal of Computer Vision (IJCV),* vol. 67, pp. 141-158, 2006.

[160] FREIBURG. (2013/12). *Freiburg Computer Vision Group*. Available: http://lmb.informatik.uni-freiburg.de/resources/binaries

[161] D. C. Jackson, G. Marmulla, M. Larinier, L. E. Miranda, and G. M. Bernacsek, "Dams, fish and fisheries. Opportunities, challenges and conflict resolution," Rome2001.

[162] S. Wu, N. Rajaratma, and C. Katopodis, "Structure of flow in vertical slot fishways," *Journal of Hydraulic Engineering,* vol. 125, pp. 351-360, 1999.

[163] J. Puertas, L. Pena, and T. Teijeiro, "An Experimental Approach to the Hydraulics of Vertical Slot Fishways," *Journal of Hydraulics Engineering,* vol. 130, 2004.

[164] L. Tarrade, A. Texier, L. David, and M. Larinier, "Topologies and measurements of turbulent flow in vertical slot fishways," *Hydrobiologia* vol. 609, pp. 177-188, 2008.

[165] H. Dewar and J. Graham, "Studies of tropical tuna swimming performance in a large water tunnel– Energetics," *Journal of Experimental Biology,* vol. 192, pp. 13-31, 1994.

[166] R. W. Blake, "Fish functional design and swimming performance," *Journal of fish biology,* vol. 65, pp. 1193-1222, 2004.

[167] J. Puertas, L. Cea, M. Bermudez, L. Pena, A. Rodriguez, J. Rabuñal*, et al.*, "Computer application for the analysis and design of vertical slot fishways in accordance with the requirements of the target species," *Ecological Engineering,* vol. 48, pp. 51-60, 2012.

[168] T. Castro-Santos, A. Haro, and S. Walk, "A passive integrated transponder (PIT) tag system for monitoring fishways," *Fisheries research,* vol. 28, pp. 253-261, 1996.

[169] J. D. Armstrong, P. M. Bagley, and I. G. Priede, "Photographic and acoustic tracking observations of the behavior of the grenadier Coryphaenoides (Nematonorus) armatus, the eel Synaphobranchus bathybius, and other abyssal demersal fish in the North Atlantic Ocean," *Marine Biology* vol. 112, pp. 1432-1793, 1992.

[170] T. W. Steig and T. K. Iverson, "Acoustic monitoring of salmonid density, target strength, and trajectories at two dams on the Columbia River, using a split-beam scaning system," *Fisheries Research,* vol. 35, pp. 43-53, 1998.

[171] R. J. Petrell, X. Shi, R. K. Ward, A. Naiberg, and C. R. Savage, "Determining fish size and swimming speed in cages and tanks using simple video techniques " *Aquacultural Engineering,* vol. 16, pp. 63-84, 1997.

[172] E. F. Morais, M. F. M. Campos, F. L. C. Padua, and R. L. Carceroni, "Particle filter-based predictive tracking for robust fish count," presented at

the Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI), 2005.

[173] S. Clausen, K. Greiner, O. Andersen, K.-A. Lie, H. Schulerud, and T. Kavli, "Automatic segmentation of overlapping fish using shape priors," presented at the Scandinavian conference on Image analysis 2007.

[174] M.-C. Chuang, J.-N. Hwang, K. Williams, and R. Towler, "Automatic fish segmentation via double local thresholding for trawl-based underwater camera systems " presented at the IEEE International Conference on Image Processing (ICIP), 2011.

[175] C. Spampinato, Y.-H. Chen-Burger, G. Nadarajan, and R. Fisher, "Detecting, Tracking and Counting Fish in Low Quality Unconstrained Underwater Videos," presented at the Int. Conf. on Computer Vision Theory and Applications (VISAPP), 2008.

[176] J. A. Lines, R. D. Tillett, L. G. Ross, D. Chan, S. Hockaday, and N. J. B. McFarlane, "An automatic image-based system for estimating the mass of free-swimming fish," *Computers and Electronics in Agriculture,* vol. 31, pp. 151-168, 2001.

[177] L. Baumgartner, M. Bettanin, J. McPherson, M. Jones, B. Zampatti, and K. Beyer, "Assessment of an infrared fish counter (Vaki Riverwatcher) to quantify fish migrations in the Murray-Darling Basin," Industry & Investment NSW AustraliaJanuary 2010 2010.

[178] V. Mitra, C.-J. Wang, and S. Banerjee, "Lidar Detection of Underwater Objects Using a Neuro-SVM-Based Architecture," *IEEE Transactions on Neural Networks,* vol. 17, pp. 717-731, 2006.

[179] A. Yilmaz, O. Javed, and M. Shah, "Object Tracking: A Survey," *ACM Computing Surveys,* vol. 38, 2006.

[180] T. Kohonen, "Self-organized formation of topologically correct feature maps," *Biol. Cybernet.,* vol. 43, pp. 59-69, 1982.

[181] A. Verikas, K. Malmqvist, and L. Bergman, "Color image segmentation by modular neural networks," *Pattern Recognition Letters* vol. 18, pp. 173-185, 1997.

[182] J. Waldemark, "An automated procedure for cluster analysis of multivariate satellite data," *Int. J. Neural Systems,* vol. 8, pp. 3-15, 1997.

[183] S.-C. Ngan and X. Hu, "Analysis of functional magnetic resonance imaging data using self-organizing mapping with spatial connectivity," *Magn. Resonance Med.,* vol. 41, pp. 939–946, 1999.

[184] M. N. Ahmed and A. A. Farag, "Two-stage neural network for volume segmentation of medical images," *Pattern Recognition Lett.,* vol. 18, pp. 1143–1151, 1997.

[185] J. Canny, "A Computational Approach To Edge Detection," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 8, pp. 679–698, 1986.

[186] M. Sezgin and B. Sankur, "Survey over image thresholding techniques and quantitative performance evaluation," *Journal of Electronic Imaging,* vol. 13, pp. 146-165, 2004.

[187] B. Stenger, P. R. S. Mendonca, and R. Cipolla, "Model-Based Hand Tracking Using an Unscented Kalman Filter," presented at the British Machine Vision Conference, 2001.

[188] K. K. C. Yu, N. R. Watson, and J. Arrillaga, "An adaptive Kalman filter for dynamic harmonic state estimation and harmonic injection tracking " *IEEE Transactions on Power Delivery,* vol. 20, pp. 1577-1584 2005.

[189] X. Yun and E. R. Bachmann, "Design, Implementation, and Experimental Results of a Quaternion-Based Kalman Filter for Human Body Motion Tracking," *IEEE Transactions on Robotics,* vol. 22, pp. 1216-1227 2006.

[190] G. Welch and G. Bishop, "An Introduction to the Kalman Filter," Department of Computer Science, University of North Carolina 2006.

[191] SIGMUR. (2003, 2013/12). *Filtering techniques. Geography degree. Tele detection*. Available: http://www.um.es/geograf/sigmur/teledet/tema06.pdf

[192] S. Heimerl, M. Hagmeyer, and C. Echteler, "Numerical flow simulation of pool-type fishways: New ways with well-known tools," *Hydrobiologia* vol. 609, pp. 189-196, 2008.

[193] L. Cea, L. Pena, J. Puertas, M. E. Vázquez-Cendón, and E. Peña, "Application of several depth-averaged turbulence models to simulate flow in vertical slot fishways," *Journal of Hydraulic Engineering,* vol. 133, pp. 160-172, 2007.

[194] J. Chorda, M. M. Maubourguet, H. Roux, M. Larinier, L. Tarrade, and L. David, "Two-dimensional free surface flow numerical model for vertical slot fishways," *Journal of Hydraulic Research,* vol. 48, pp. 141-151, 2010.

[195] M. Bermúdez, J. Puertas, L. Cea, L. Pena, and L. Balairón, "Influence of pool geometry on the biological efficiency of vertical slot fishways," *Ecological Engineering,* vol. 36, pp. 1355-1364, 10// 2010.

Anexo: Resumen de la Tesis en Castellano

Hoy en día los sistemas inteligentes son una de las mejores aproximaciones para tratar problemas relacionados con la optimización, la imprecisión y la incertidumbre.

Mientras estas nuevas técnicas evolucionan, la manera en que las nuevas tecnologías interaccionan con el mundo real está también cambiando rápidamente, y la percepción a través de la visión es el nuevo paradigma de un nuevo tipo de computación y de robótica, donde las maquinas podrían por primera vez manejar datos y representaciones del mundo comprensibles por los humanos.

Usando este nivel compartido de percepción, aplicaciones basadas en imágenes enfocadas al usuario medio y para el uso cotidiano, están emergiendo en un mundo donde los dispositivos fotográficos personales son un nuevo estándar social.

La Visión Artificial (*VA*), el campo enfocado a la extracción de conocimiento de las imágenes utilizando computadoras, está en el centro de esta revolución tecnológica.

En Ingeniería Civil, y otras áreas como la biomedicina, donde es extremadamente importante medir procesos observables, alcanzando resultados precisos y a un bajo coste, este tipo de técnicas proporcionan una nueva aproximación para resolver problemas tradicionales con un grado de precisión y flexibilidad nunca antes conocido y para enfrentarse a nuevos desafíos.

Desde el punto de vista de los sistemas de información, el propósito de la visión no se basa solo en la obtención de datos con los que describir una escena, sino que debe considerarse como un proceso activo donde, para alcanzar un objetivo particular, es necesario utilizar información, representaciones y algoritmos específicos.

En este dominio, el objetivo de un sistema de *VA* está bien definido y el trabajo del diseñador radica en encontrar todos los componentes y mecanismos para alcanzar dicho propósito.

Actualmente, debe aceptarse que no existe un algoritmo general de *VA*, siendo necesario desarrollar un sistema específico enfocado al problema concreto que se desea resolver.

Teniendo en cuenta esta situación, el objetivo principal de esta tesis será el de definir una metodología para desarrollar sistemas de *VA* aplicados a Ingeniería Civil, atendiendo a las características específicas de esta área.

Para elaborar un sistema de Visión Artificial utilizando una aproximación sistemática, el primer paso es elaborar una descripción precisa del sistema,

describiendo las técnicas aplicaciones y fundamentos teóricos generales de la disciplina. Identificando los procesos, los componentes y relaciones típicas entre ellos, y analizando los requisitos de los problemas concretos a los que se enfrentará el diseñador.

La Ingeniería Civil es una disciplina que trata sobre el diseño, la construcción y el mantenimiento de diversas estructuras como carreteras, puentes, canales o embalses entre otras, así como de su relación con el medio.

La Ingeniería Civil está dividida en diferentes subáreas, como la geodésica, la geofísica, la ingeniería de materiales, la ingeniería del transporte o la hidrología.

En esta disciplina, los ingenieros civiles realizan una amplia gama de experimentos con el fin de evaluar y comparar diferentes configuraciones de diseño, analizar las propiedades de diversos materiales o seleccionar parámetros de diseño.

De este modo, un experimento puede definirse básicamente como un test o un conjunto de tests en los cuales se realizan cambios en una o una serie de variables de entrada para observar la respuesta obtenida, midiendo las variaciones en uno o más parámetros de salida.

En Ingeniería Civil, normalmente, estos experimentos no pueden realizarse directamente en los sistemas reales que desean estudiarse, porque dichos sistemas no están disponibles, porque los ensayos son destructivos o simplemente porque en los sistemas reales no pueden controlarse las diferentes condiciones experimentales necesarias para obtener conclusiones válidas.

En estos casos, una de las posibles alternativas es la de realizar una simulación en computador del sistema deseado, basándose en la elaboración de un modelo

numérico o computacional. Sin embargo, en muchos casos, los procesos a estudiar son demasiado complejos para ser comprendidos analíticamente o los modelos existentes no han sido confirmados en el mundo real.

En este tipo de escenarios, los ingenieros civiles realizan sus experimentos en modelos a escala. Es decir, modelos físicos de tamaño reducido que mantienen las propiedades del sistema de interés.

En estos experimentos, los más comunes en Ingeniería Civil, es necesario realizar diferentes mediciones de magnitudes físicas que determinarán la respuesta del sistema a la variación de los parámetros de entrada.

Con este objetivo, en Ingeniería Civil se utilizan instrumentos de medición específicos, como galgas extensiométricas para medir la deformación entre dos puntos de un material, o caudalímetros, para cuantificar el flujo hidráulico.

La mayoría de estos instrumentos son caros y proporcionan mediciones muy específicas, restringidas a los puntos en que son colocados, e interfirieren físicamente con el proceso que se desea medir.

La Visión Artificial, permite obtener mediciones precisas de prácticamente cualquier fenómeno físico observable, proporcionando una alternativa más flexible, precisa y barata que muchos de los instrumentos tradicionales.

Esta es la hipótesis de partida de la presente tesis, donde se analizará la aplicación sistemática de las técnicas y métodos de la *VA* para obtener una metodología que permita el desarrollo planificado de un sistema genérico aplicado a Ingeniería Civil.

Con este objetivo, en la presente tesis se realiza un análisis de las estrategias y principios generales de la disciplina, teniendo en cuenta un escenario de aplicación

práctica con una perspectiva similar a los sistemas industriales; empezando por la identificación y descomposición de cada problema en tareas generales, exponiendo a continuación las diferentes técnicas que permiten abordar su análisis y finalmente definiendo los componentes del sistema y las diferentes etapas del ciclo de vida del mismo.

Para validar la hipótesis de partida, en esta tesis se expondrán además, dos sistemas diferentes, elaborados según la metodología propuesta, para su aplicación en dos escenarios de Ingeniería Civil.

1 Estructura de la Tesis

Esta tesis está dividida en los siguientes capítulos:

- En el capítulo 1, se exponen la motivación, objetivos y contribuciones generales de las investigaciones realizadas.
- En el capítulo 2, se analizan en profundidad los fundamentos de la disciplina y los sistemas de Visión Artificial desde la perspectiva de la Ingeniería Civil.
- En el capítulo 3, se expone la metodología propuesta, categorizando los elementos y etapas característicos de un sistema de *VA*, así como las relaciones entre ellos, para abordar posteriormente los pasos que deben seguirse para diseñar y construir un sistema de forma planificada y sistemática.
- En el capítulo 4, se expondrá el primero de los sistemas desarrollados con la metodología propuesta, cuyo objetivo es el de analizar ópticamente ensayos de deformación de materiales, obteniendo información del desplazamiento y la deformación del material a lo largo del tiempo y en cada punto. De este modo se proporciona una alternativa a las herramientas habituales en Ingeniería Civil

para este propósito, como las galgas extensiométricas y los extensómetros de contacto.

- En el capítulo 5, se presenta el segundo sistema de *VA* desarrollado con la metodología propuesta, cuyo objetivo es el de analizar la trayectoria y el comportamiento de peces en el interior de estructuras hidráulicas conocidas como escalas de hendidura vertical.

2 Metodología Propuesta

El propósito de una metodología es el de establecer un marco de trabajo para crear o mantener un sistema de forma deliberada y metódica. Ello no solo reduce el tiempo y coste del desarrollo de un sistema, sino que también hará más sencilla la tarea de su diseño y construcción, aumentando la eficiencia del trabajo y la calidad del producto final.

El objetivo final de la metodología propuesta es el de realizar un sistema para medir un proceso físico de interés que tendrá lugar en condiciones de laboratorio controladas en el marco de la Ingeniería Civil.

Usando un análisis intuitivo, el primer paso de la metodología es el de estudiar el proceso de interés así como las diferentes aplicaciones de *VA* que se han aplicado con éxito en procesos similares.

Posteriormente, se definen los elementos necesarios y el flujo de trabajo desde el dominio del problema hasta la obtención de la solución final.

Un sistema de VA, aplicado en este escenario, se compondrá como norma general de los mismos elementos y etapas que un sistema de inspección industrial:

- **Adquisición de la imagen:** Un dispositivo físico sensible a una banda del espectro electromagnético que proporcionará un registro digital de la escena de interés.
- **Almacenamiento:** Donde se almacenarán las imágenes e información adquiridas en la etapa anterior.
- **Procesado:** El elemento donde las imágenes serán analizadas, extrayendo la información necesaria para tomar las decisiones que sean requeridas.
- **Comunicación:** Los elementos que se encargarán de del acceso a los datos desde diferentes dispositivos.
- **Visualización:** Típicamente formado por monitores u otros elementos que permitirán el control y la supervisión del proceso.
- **Actuadores:** Donde se utilizará de forma reactiva la información extraída y procesada por el sistema para realizar cualquier tipo de acción como el control dinámico del sistema o cualquier acción sobre el proceso estudiado.

Así mismo, las etapas de procesado de un sistema de *VA* seguirán por lo general el mismo flujo de trabajo:

- **Procesado de bajo nivel:** Las imágenes se mejoran y se preparan para las siguientes etapas. No se realiza una interpretación de la información.
- **Procesado intermedio:** Se aplican técnicas de análisis basadas en conocimiento *a priori* e hipótesis sencillas.
- **Procesado de alto nivel:** En este nivel se realizan procesos más complejos que requieren una interpretación de la información de la imagen.

Aunque los diferentes problemas a resolver impondrán el uso de diferentes técnicas o algoritmos, se pueden definir de forma general las tareas requeridas en un sistema genérico:

- **Adquisición de la imagen**
- **Preprocesado:** Se utilizarán técnicas aplicadas a la imagen para realzar la información útil o reducir la información no deseada.
- **Segmentación:** Consiste en la separación de los diferentes objetos o estructuras que componen la imagen. En esta etapa se utiliza conocimiento sencillo obtenido *a priori*, y a menudo basado en hipótesis sobre las condiciones experimentales.
- **Representación y descripción:** Los objetos detectados serán traducidos a una forma útil y significativa con la que se trabajará en las siguientes etapas.
- **Reconocimiento e interpretación:** Las representaciones de los diferentes objetos se interpretan, asociándolas a conceptos o clases del mundo real. En esta etapa se requiere la utilización de diferentes tipos de conocimiento, ya sea información sobre los objetos reales representados en la imagen o información obtenida por el propio sistema.
- **Análisis de alto nivel:** Esta etapa involucra la utilización de los modelos elaborados en la fase de interpretación para sintetizar conocimiento útil.
- **Análisis de resultados y toma de decisiones:** En esta etapa se procesan los resultados, obteniendo conclusiones y permitiendo la toma de acciones basadas en dichas conclusiones.

Una vez identificadas las etapas, elementos y el conocimiento necesarios, se seguirá una metodología de desarrollo donde se contemplará el ciclo de vida del sistema y en la que se seguirá una aproximación basada en un esquema mixto de prototipado y desarrollo en espiral y donde se definen con detalle las diferentes fases involucradas, desde la investigación inicial hasta la obtención del producto final.

Posteriormente, se definen estrategias típicas de calibración de la cámara utilizando el modelo proyectivo de pin-hole, extendido con los modelos numéricos de las distorsiones más típicas producidas por la lente.

Finalmente, se tratan las diferentes técnicas de iluminación que permiten resaltar diferentes características de los objetos estudiados como las texturas y la elevación.

3 Aplicación para el Análisis de Materiales

Uno de los problemas más importantes en Ingeniería Civil es el análisis del comportamiento de materiales en ensayos de resistencia donde típicamente se miden los desplazamientos y deformaciones de los materiales sometidos a diferentes esfuerzos.

Estas mediciones se realizan tradicionalmente con dispositivos como las galgas extensiométricas, que están limitadas a realizar mediciones en un punto concreto del material donde son colocadas, y en una dirección determinada.

En esta tesis se presenta un sistema para la medición del desplazamiento y la deformación del material basado en el análisis de las grabaciones del ensayo de resistencia.

La técnica propuesta se basa en la integración de un algoritmo de calibración con una técnica de medición de desplazamiento conocida como *Block-Matching*. Esta técnica se basa en la correlación de diferentes regiones de la imagen, denominadas bloques, con regiones candidatas en las siguientes imágenes de la secuencia.

En el algoritmo de *Block-Matching*, cada región o bloque candidato es asociada a un vector que corresponde a su desplazamiento estimado. En el algoritmo propuesto, se utiliza una aproximación iterativa donde tienen lugar las siguientes fases:

- **Definición de los bloques:** Las regiones iniciales donde se efectuará el análisis se definen teniendo en cuenta la forma de los diferentes objetos, calculados mediante una etapa de segmentación y alrededor de determinados puntos de interés, determinados por el usuario o a través de una técnica de análisis de características.
- **Deformación:** El área de búsqueda y la forma de los bloques candidatos se actualizan usando el desplazamiento calculado en la iteración anterior.
- **Descomposición piramidal:** Se trata de una etapa de análisis multiresolución donde la imagen se remuestrea, reduciendo su tamaño, para encontrar el mejor valor global.
- **Medición de similitud:** Se utiliza una medida numérica para calcular el mejor candidato discreto para cada bloque en la imagen objetivo.
- **Filtrado y suavizado de los vectores:** Se eliminan los vectores anómalos y se garantizan unos resultados más homogéneos utilizando una media ponderada de los vecinos de cada vector.
- **Modelo de deformación:** Los vectores encontrados para cada bloque se utilizan para encontrar el desplazamiento de la imagen en cada punto de acuerdo a un modelo matemático de deformación que se utilizará para interpolar y restringir los grados de libertad de los desplazamientos encontrados.

La técnica propuesta se ha validado en diferentes ensayos, comparando su precisión con otras técnicas de imagen e instrumentos de Ingeniería Civil en diferentes ensayos con imágenes sintéticas y con ensayos de deformación reales.

Los resultados obtenidos en los ensayos mejoran en la mayoría de los casos al de los obtenidos con otras técnicas de *VA* y son comparables a los obtenidos con la instrumentación usada en Ingeniería Civil, obteniéndose mediciones más flexibles y económicas.

4 Aplicación para Seguimiento de Peces

Las escalas de hendidura vertical son estructuras hidráulicas que garantizan a la fauna autóctona el movimiento a través del río, sorteando las construcciones transversales, como embalses o centrales hidroeléctricas que obstruyen el cauce natural del río.

Estas estructuras se diseñan de acuerdo a sus propiedades hidráulicas, pero en su diseño no se tienen en cuenta los requisitos biológicos de los peces ya que su comportamiento en el interior de la escala no puede ser estudiado adecuadamente con los medios de los que se dispone actualmente.

En esta tesis se propone un sistema para medir la trayectoria y el comportamiento de los peces dentro de estas estructuras, basado en el análisis de las imágenes recogidas en ensayos con peces reales.

Con este objetivo, se ha diseñado un sistema de adquisición de imágenes compuesto por una serie de cámaras cenitales parcialmente sumergidas en el agua y diferentes dispositivos de almacenamiento y visualización.

El sistema desarrollado se acopla a una escala donde se realizan una serie de ensayos con peces. Las imágenes obtenidas son analizadas con un algoritmo de VA.

El algoritmo propuesto se compone de las siguientes etapas:

- **Calibración de la cámara:** Se ha realizado un algoritmo para eliminar la distorsión de la imagen y proyectar las mediciones obtenidas de cada cámara en un modelo de coordenadas común que tiene en cuenta el solapamiento entre las diferentes cámaras y la refracción provocada por el agua.
- **Segmentación:** El fondo de la imagen es separado de las regiones donde se encuentra el pez. Con este fin se combinan diferentes técnicas basadas en el análisis de la discontinuidad de la imagen (análisis de bordes) y el análisis de las diferentes regiones de la misma.
- **Representación e interpretación:** Los objetos detectados son traducidos a una representación descriptiva con la que se operará en las siguientes etapas del proceso. En esta etapa se realiza además un primer filtrado de los resultados.
- **Seguimiento o *tracking*:** Los objetos detectados se utilizan como entrada de un filtro basado en el modelo de Kalman. Este es un algoritmo que utiliza un modelo de desplazamiento y que actúa como un modelo bayesiano, permitiendo: la eliminación de detecciones aisladas provocadas por el ruido, la separación de las trayectorias de los diferentes individuos, la estimación de las posiciones donde los peces han estado ocultos y la predicción de donde se encontrarán los peces en el futuro.
- **Análisis de los datos:** Las posiciones de los peces en la escala se proyectan sobre un mapa de velocidades del agua calculado con modelos numéricos. De esta forma, a partir de la velocidad observada se puede medir la velocidad real de natación en cada punto y calcular diferentes medidas sobre el comportamiento de los peces.

Después de analizar la precisión de la técnica propuesta, el sistema desarrollado se ha validado con numerosos ensayos donde se ha obtenido información sobre el ascenso a través de la escala de más de 200 peces de cuatro especies diferentes.

La información recogida ha permitido analizar las velocidades y aceleraciones medias y máximas de los peces, sus tiempos de descanso y las zonas de la escala donde se han efectuado estos descansos, así como encontrar sus trayectorias típicas y medir su efectividad en el ascenso de la escala.

Los resultados obtenidos en estos ensayos han sido enormemente prometedores, y se emplearán en mejorar la metodología de diseño de las escalas y en mejorar la evaluación del rendimiento de las escalas existentes.

5 Conclusiones y Desarrollos Futuros

Esta tesis se ha enfocado en la aplicación de técnicas de Visión Artificial en el campo de la Ingeniería Civil, bajo la hipótesis de que estas técnicas pueden constituir una importante alternativa a algunos de los instrumentos de medición utilizados tradicionalmente en este área.

Para validar esta hipótesis, se ha desarrollado una metodología para el diseño de estos sistemas, orientada a los requisitos específicos de la Ingeniería Civil.

De acuerdo con esta metodología, se han diseñado dos sistemas de *VA* para el análisis de ensayos de materiales y para el seguimiento de peces en escalas de hendidura vertical.

Los resultados obtenidos con ambos sistemas representan valiosas contribuciones en sus diferentes campos y han sido publicados en varias revistas de prestigio internacional.

De este modo, puede concluirse que la metodología propuesta es útil y se valida la hipótesis de partida de esta tesis.

En futuros desarrollos de este trabajo, la metodología propuesta se ampliará para proporcionar mejores y más precisas guías de actuación y diseño para la utilización de sistemas de Visión Artificial en Ingeniería Civil.

Además, se espera que la metodología propuesta sea aplicada en el futuro a nuevos escenarios y que sea extendida para incorporar mediciones en 3 dimensiones y nuevos dispositivos ópticos.

Anexo: Resumo da Tese en Galego

Hoxe en día os sistemas intelixentes son unha das mellores aproximacións para tratar problemas relacionados ca optimización, a imprecisión e a incerteza.

Mentres estas novas técnicas evolucionan, a maneira en que as novas tecnoloxías interaccionan co mundo real está tamén cambiando rapidamente, e a percepción a través da visión é o novo paradigma dun tipo de computación e de robótica onde as maquinas poderían, por primeira vez manexar datos e representacións do mundo comprensibles polos humanos.

Usando este nivel compartido de percepción, aplicacións baseadas en imaxes, enfocadas o usuario medio e para o o uso cotiá, están a emerxer en un mundo onde os dispositivos fotográficos persoais son un novo estándar social.

A Visión Artificial (*VA*), o campo enfocado á extracción de coñecemento das imaxes usando computadoras, está no centro de esta revolución tecnolóxica.

En Enxeñería Civil, e outras areas como a biomedicina, onde é extremadamente importante medir procesos observables, acadando resultados precisos e a un baixo custo, este tipo de técnicas proporcionan unha nova aproximación para resolver problemas tradicionais con un grado de precisión e flexibilidade nunca antes coñecido e para enfrontarse a novos desafíos.

Dende o punto de vista dos sistemas de información, o propósito da visión non se basea so na obtención de datos cos que describir unha escena, senón que debe considerarse como un proceso activo, onde, para acadar un obxectivo particular é necesario utilizar información, representacións e algoritmos específicos.

En este dominio, o obxectivo de un sistema de *VA* está ben definido e o traballo do deseñador radica en atopar todos os compoñentes e mecanismos para acadar dito propósito.

Na actualidade debe aceptarse que non existe un algoritmo xeral de *VA,* sendo necesario desenvolver un sistema específico enfocado ó problema concreto que se desexa resolver.

Tendo en conta esta situación, o obxectivo principal de esta tese será o de definir unha metodoloxía para desenvolver sistemas de *VA* aplicados á Enxeñería Civil, atendendo ás características específicas de esta área.

Para elaborar un sistema de Visión Artificial utilizando unha aproximación sistemática. O primeiro paso e elaborar unha descrición precisa do sistema, describindo as técnicas aplicacións e fundamentos teóricos xerais da disciplina.

Identificando os procesos, os compoñentes e as relacións entre eles e analizando os requisitos dos problemas concretos ós que se enfrontará o deseñador.

A Enxeñería Civil e unha disciplina que trata sobre o deseño, a construción e o mantemento de diversas estruturas como carreteiras, pontes, canles ou encoros entre outras, así como da súa relación co medio.

A Enxeñería Civil está dividida en diferentes subáreas como a xeodésica, a xeofísica, a enxeñería de materiais, a enxeñería do transporte ou a hidroloxía.

En esta disciplina, os enxeñeiros civís realizan unha ampla gama de experimentos co fin de avaliar e comparar diferentes configuracións de deseño, analizar as propiedades de diversos materiais ou seleccionar parámetros de deseño.

De este modo, un experimento pode definirse basicamente coma un test ou un conxunto de tests nos cales se realizan cambios en unha ou unha serie de variables de entrada para observar a resposta obtida, medindo as variacións en un ou máis parámetros de saída.

En Enxeñería Civil, normalmente, estes experimentos non poden realizarse directamente sobre os sistemas reais que desexan estudarse, porque ditos sistemas non están dispoñibles, porque os ensaios son destrutivos ou simplemente porque nos sistemas reais non poden controlarse as diferentes condicións experimentais necesarias para obter conclusións válidas.

En estes casos, unha das posibles alternativas é a de realizar unha simulación por computador do sistema desexado, baseándose na elaboración dun modelo numérico ou computacional. Sen embargo, en moitos casos, os procesos a estudar son demasiado complexos para ser comprendidos analiticamente ou os modelos existentes non foron confirmados no mundo real.

En este tipo de escenarios, os enxeñeiros civís realizan os seus experimentos en modelos a escala. É dicir, modelos físicos de tamaño reducido que manteñen as propiedades do sistema de interese.

En estes experimentos, os máis comúns en Enxeñería Civil, é necesario realizar diferentes medicións de magnitudes físicas que determinan a resposta do sistema fronte á variación dos parámetros de entrada.

Con este obxectivo, en Enxeñería Civil utilízanse instrumentos de medición específicos como as galgas extensiométricas para medir a deformación entre dous puntos de un material ou caudalímetros, para cuantificar o fluxo hidráulico.

A maioría destes instrumentos son caros e proporcionan medicións moi específicas, restrinxidas ós puntos en que son colocados, e interfiren fisicamente co proceso que se desexa medir.

A Visión Artificial permite obter medicións precisas de practicamente calquera fenómeno físico observable, proporcionando unha alternativa máis flexible, precisa e barata que moitos dos instrumentos tradicionais.

Esta é a hipótese de partida da presente tese, onde se analizará a aplicación sistemática das técnicas e métodos da *VA* para obter unha metodoloxía que permita o desenvolvemento planificado de un sistema xenérico aplicado á Enxeñería Civil.

Con este obxectivo, na presente tese realizase un análise das estratexias e principios xerais da disciplina, tendo en conta un escenario de aplicación practica e cunha perspectiva similar á dos sistemas industriais; empezando pola identificación e descomposición de cada problema en tarefas xerais, expoñendo a continuación as diferentes técnicas que permitan abordar o seu análise e

finalmente definindo as compoñentes do sistema e as diferentes etapas do ciclo de vida do mesmo.

Para validar a hipótese de partida, en esta tese exporanse, ademais, dous sistemas diferentes, elaborados segundo a metodoloxía proposta, para a súa aplicación en dous escenarios de Enxeñería Civil.

1 Estrutura da Tese

Esta tese está dividida nos seguintes capítulos:

- No capítulo 1, expóñense a motivación, obxectivos e contribucións xerais das investigacións realizadas.
- No capítulo 2, analizanse en profundidade os fundamentos da disciplina e os sistemas de Visión Artificial desde a perspectiva da Enxeñería Civil.
- No capítulo 3, exponse a metodoloxía proposta, categorizando os elementos e etapas característicos de un sistema de *VA*, así como as relaciones entre eles, para abordar posteriormente os pasos que deben seguirse para deseñar e construír un sistema de forma planificada e sistemática.
- No capítulo 4, exponse o primeiro dos sistemas desenvolvidos ca metodoloxía proposta. O seu obxectivo é o de analizar ópticamente ensaios de deformación de materiais, obtendo información do desprazamento e la deformación do material ó longo do tempo e en cada punto. De este modo proporcionase unha alternativa ás ferramentas habituais en Enxeñería Civil para este propósito, como as galgas extensiométricas e los extensómetros de contacto.
- No capítulo 5, presentase o segundo sistema de *VA* desenvolto ca metodoloxía proposta. O seu obxectivo é o de analizar a traxectoria e o comportamento de

peixes no interior de estruturas hidráulicas coñecidas como escalas de fenda vertical.

2 Metodoloxía Proposta

O propósito de unha metodoloxía é el de establecer un marco de traballo para crear ou manter un sistema de forma deliberada e metódica. Elo non so reduce o tempo e custo de desenvolvemento dun sistema, senón que tamén fará máis sinxela a tarefa do seu deseño e construción, aumentando a eficiencia do traballo e a calidade do produto final.

O obxectivo final da metodoloxía proposta é o de realizar un sistema para medir un proceso físico de interese que terá lugar en condicións de laboratorio controladas no marco da Enxeñería Civil.

Usando un análise intuitivo, o primeiro paso da metodoloxía é o de estudar o proceso de interese así como las diferentes aplicacións de *VA* que se aplicaron con éxito en procesos similares.

Posteriormente, se definen los elementos necesarios y el flujo de trabajo desde el dominio del problema hasta la obtención de la solución final.

Un sistema de VA, aplicado en este escenario, comporase por norma xeral dos mesmos elementos e etapas que un sistema de inspección industrial:

- **Adquisición da imaxe:** Un dispositivo físico sensible a unha banda do espectro electromagnético que proporcionará un rexistro dixital da escena de interese.

- **Almacenamento:** Onde se almacenarán as imaxes e información adquiridas na etapa anterior.
- **Procesado:** O elemento onde as imaxes serán analizadas, extraendo a información necesaria para tomar as decisións que sexan requiridas.
- **Comunicación:** Os elementos que se encargarán do acceso ós datos dende diferentes dispositivos.
- **Visualización:** Tipicamente formado por monitores ou outros elementos que permitirán o control e a supervisión do proceso.
- **Actuadores:** Onde se utilizará de forma reactiva a información extraída e procesada polo sistema para realizar calquera tipo de acción como o control dinámico do sistema o calquera acción sobre o proceso estudado.

Así mesmo, as etapas de procesado dun sistema de *VA* seguirán polo xeral o mesmo fluxo de traballo:

- **Procesado de baixo nivel:** As imaxes mellóranse e prepáranse para as seguintes etapas. Non se realiza una interpretación da información.
- **Procesado intermedio:** Aplícanse técnicas de análise baseadas no coñecemento *a priori* e hipóteses sinxelas.
- **Procesado de alto nivel:** En este nivel realízanse procesos máis complexos que requiren unha interpretación da información da imaxe.

Aínda que os diferentes problemas a resolver imporán o uso de diferentes técnicas ou algoritmos, pódense definir de forma xeral as tarefas requiridas nun sistema xenérico:

- **Adquisición da imaxe**
- **Preprocesado:** Utilízanse técnicas aplicadas á imaxe para realzar a información útil ou reducir a información non desexada.

- **Segmentación:** Consiste na separación dos diferentes obxectos ou estruturas que compoñen a imaxe. En esta etapa utilizase coñecemento sinxelo obtido *a priori*, e a miúdo baseado en hipóteses sobre as condicións experimentais.
- **Representación e descrición:** Os obxectos detectados son traducidos a unha forma útil e significativa ca que se traballará nas seguintes etapas.
- **Recoñecemento e interpretación:** As representacións dos diferentes obxectos interprétanse, asociándoas a conceptos ou clases do mundo real. En esta etapa requírese a utilización de diferentes tipos de coñecemento, xa sexa información sobre os obxectos reais representados na imaxe ou información obtida polo propio sistema.
- **Análise de alto nivel:** Esta etapa involucra a utilización dos modelos elaborados na fase de interpretación para sintetizar coñecemento útil.
- **Análise de resultados e toma de decisións:** En esta etapa procésanse os resultados, obtendo conclusións e permitindo a toma de accións baseadas en ditas conclusións.

Unha vez identificadas as etapas, os elementos e o coñecemento necesarios, seguirase unha metodoloxía de desenvolvemento onde se contemplará o ciclo de vida do sistema e na que se seguirá unha aproximación baseada en un esquema mixto de prototipado e desenvolvemento en espiral y onde se definen con detalle as diferentes fases involucradas, desde a investigación inicial ata la obtención do produto final.

Posteriormente, defínense estratexias típicas de calibración dea cámara utilizando o modelo proxectivo de pin-hole, estendido cos modelos numéricos das distorsións máis típicas producidas pola lente.

Finalmente, trátanse as diferentes técnicas de iluminación que permiten resaltar diferentes características dos obxectos estudados como as texturas e la elevación.

3 Aplicación para o Análise de Materiais

Un dos problemas máis importantes en Enxeñería Civil é o análise do comportamento de materiais en ensaios de resistencia onde tipicamente se miden os desprazamentos e deformacións dos materiais sometidos a diferentes esforzos.

Estas medicións realízanse tradicionalmente con dispositivos como as galgas extensiométricas, que están limitadas a realizar medicións en un punto concreto do material onde son colocadas, e nunha dirección determinada.

En esta tese presentase un sistema para a medición do desprazamento e a deformación do material baseado na análise das gravacións do ensaio de resistencia.

A técnica proposta basease na integración dun algoritmo de calibración cunha técnica de medición de desprazamento coñecida como *Block-Matching*. Esta técnica basease na correlación de diferentes rexións da imaxe, denominadas bloques, con rexións candidatas nas seguintes imaxes da secuencia.

No algoritmo de *Block-Matching*, cada rexión ou bloque candidato é asociada a un vector que se corresponde co seu desprazamento estimado. No algoritmo proposto, utilizase una aproximación iterativa onde teñen lugar as seguintes fases:

- **Definición dos bloques:** As rexións iniciais onde se efectuará o análise defínense tendo en conta a forma dos diferentes obxectos, calculados mediante unha etapa de segmentación e ó redor de determinados puntos de interese, determinados polo usuario ou a través dunha técnica de análise de características.
- **Deformación:** A área de busca e la forma dos bloques candidatos actualízanse usando o desprazamento calculado na iteración anterior.

- **Descomposición piramidal:** Tratase dunha etapa de análise multiresolución onde a imaxe se remuestrea, reducindo o seu tamaño, para encontrar o mellor valor global.
- **Medición de similitude:** Utilízase unha medida numérica para calcular o mellor candidato discreto para cada bloque na imaxe obxectivo.
- **Filtrado e suavizado dos vectores:** Elimínanse os vectores anómalos e garántense uns resultados máis homoxéneos utilizando unha media ponderada dos veciños de cada vector.
- **Modelo de deformación:** Os vectores encontrados para cada bloque utilízanse para atopar o desprazamento da imaxe en cada punto de acordo a un modelo matemático de deformación que se utilizará para interpolar e restrinxir os grados de liberdade dos desprazamentos atopados.

A técnica proposta validouse en diferentes ensaios, comparando a súa precisión con outras técnicas de imaxe e instrumentos de Enxeñería Civil en diferentes ensaios con imaxes sintéticas e con ensaios de deformación reais.

Os resultados obtidos nos ensaios melloran na maioría de los casos ó dos obtidos con outras técnicas de *VA* e son comparables ós obtidos ca instrumentación usada en Enxeñería Civil, obténdose medicións máis flexibles e económicas.

4 Aplicación para o Seguimento de Peixes

As escalas de fenda vertical son estruturas hidráulicas que garanten á fauna autóctona o movemento a través do río, sorteando as construcións transversais, como encoros ou centrais hidroeléctricas que obstrúen o curso natural do río.

Estas estruturas deséñanse de acordo ás súas propiedades hidráulicas, pero no seu deseño non se teñen en conta os requisitos biolóxicos dos peixes xa que o seu comportamento no interior de la escala non pode ser estudado adecuadamente cos medios dos que se dispón actualmente.

En esta tese proponse un sistema para medir a traxectoria e eo comportamento dos peixes dentro de estas estruturas, baseado na análise das imaxes recollidas en ensaios con peixes reais.

Con este obxectivo, deseñouse un sistema de adquisición de imaxes composto por unha serie de cámaras cenitais parcialmente somerxidas na auga e diferentes dispositivos de almacenamento e visualización.

O sistema acóplase a unha escala onde se realizan unha serie de ensaios con peixes. As imaxes obtidas son analizadas cun algoritmo de VA.

O algoritmo proposto componse das seguintes etapas:

- **Calibración da cámara:** Realizouse un algoritmo para eliminar a distorsión da imaxe e proxectar as medicións obtidas de cada cámara nun modelo de coordenadas común que ten en conta o solapamento entre as diferentes cámaras e la refracción provocada pola auga.
- **Segmentación:** O fondo da imaxe separase das rexións onde se atopa o peixe. Con este fin combínanse diferentes técnicas baseadas na análise da descontinuidade da imaxe (análise de bordes) e o análise das diferentes rexións da mesma.
- **Representación e interpretación:** Os obxectos detectados son traducidos a unha representación descritiva ca que se opera nas seguintes etapas do proceso. En esta etapa realizase ademais un primeiro filtrado deos resultados.

- **Seguimento ou *tracking*:** Os obxectos detectados utilízanse como entrada dun filtro baseado no modelo de Kalman. Este é un algoritmo que utiliza un modelo de desprazamento e que actúa como un modelo bayesiano, permitindo: a eliminación de detecciónsilladas provocadas polo ruído, a separación das traxectorias dos diferentes individuos, a estimación das posicións onde os peixes estiveron ocultos e a predición de onde se encontrarán os peixes no futuro.
- **Análise dos datos:** As posicións dos peixes na escala proxéctanse sobre un mapa de velocidades da auga calculado con modelos numéricos. Desta forma, a partir da velocidade observada pódese medir a velocidade real de natación en cada punto e calcular diferentes medidas sobre o comportamento dos peixes.

Despois de analizar a precisión da técnica creada, o sistema proposto validouse con numerosos ensaios onde se obtivo información sobre o ascenso a través da escala de máis de 200 peces de catro especies diferentes.

La información recollida permitiu analizar as velocidades e aceleracións medias e máximas dos peixes, os seus tempos de descanso e as zonas da escala onde efectuaron estes descansos, así como encontrar as súas traxectorias típicas e medir a súa efectividade no ascenso da escala.

Os resultados obtidos en estes ensaios foron enormemente prometedores, e empregaranse en mellorar a metodoloxía de deseño das escalas e en mellorar a avaliación do rendemento das escalas existentes.

5 Conclusións e Traballo Futuro

Esta tese enfocouse na aplicación de técnicas de Visión Artificial no campo da Enxeñería Civil, baixo a hipótese de que estas técnicas poden constituír unha importante alternativa a algúns dos instrumentos de medición utilizados tradicionalmente neste área.

Para validar esta hipótese, desenvolveuse unha metodoloxía para o deseño destes sistemas, orientada ós requisitos específicos da Enxeñería Civil.

De acordo con esta metodoloxía, deseñáronse dous sistemas de *VA* para a análise de ensaios de materiais e para o seguimento de peixes en escalas de fenda vertical.

Os resultados obtidos con ambos sistemas representan valiosas contribucións nos seus diferentes campos e foron publicados en varias revistas de prestixio internacional.

De este modo, pode concluírse que a metodoloxía proposta é útil e validase a hipótese de partida de esta tese.

En futuros traballos, a metodoloxía proposta ampliarase para proporcionar mellores e más precisas guías de actuación e deseño para a utilización de sistemas de Visión Artificial en Enxeñería Civil.

Ademais, esperase que a metodoloxía proposta sexa aplicada no futuro a novos escenarios e que sexa estendida para incorporar medicións en 3 dimensións e novos dispositivos ópticos.

But I, being poor, have only my dreams;
I have spread my dreams under your feet;
Tread softly because you tread on my dreams.
William Butler Yeats