# OBJECT FLOW
## A per-object dense motion descriptor

Juan Manuel Pérez Rúa

technicolor

UNIVERSITÉ DE BOURGOGNE

Technicolor SA - University of Burgundy

A Thesis Submitted for the Degree of
MSc Erasmus Mundus in Vision and Robotics (VIBOT)

· 2014 ·

**Abstract**

Recent advances in motion analysis in image sequences have shown undoubtedly impressive progresses. Among them, dense (i.e. for every pixel of the image) optical flow estimation and per-object visual tracking are fundamental still-open problems and, as independent bricks in the computer vision toolbox, high accuracy algorithms are at hand. However, combining such challenges in a unified framework solution remains an unexplored terrain. Thus, the present work aims towards the posibility of exploiting robust object tracking algorithms for boosting per pixel motion estimation. And, subsequently, to find out if a semantic (high-level) decomposition of the scene together with a (low-level) motion characterization, would permit advanced operations in a per-object basis.

*Todo arde si le aplicas la chispa adecuada....*

Enrique Bunbury

# Contents

# List of Figures

# List of Tables

# Acknowledgments

...

# Chapter 1

# Introduction

## 1.1 Preparing your dissertation

You are strongly encouraged to use the Latex templates provided.

### 1.1.1 Paper

The manuscript should be in A4 size, and the printed paper should be of at least 70 gsm.

### 1.1.2 Font and margins

Thesis should be printed on both sides of the paper. Use no less than 1.5 spacing, with quotations and notes single-spaced. Regarding **Character size**, not less than 2.0mm for capitals and 1.5mm for x-height (the height of a lower-case x). Us a serif font (i.e. Times) between 10 and 12 points. Use consistent and clear fonts through all the document.

The text layout should be approximately as follows:

- $4cm$ binding margin

- $2cm$ head margin (top of page)

- $2.5cm$ fore-edge margin

- $4cm$ tail margin (bottom of page)

## 1.2 Title Page

The title page should contain the title of thesis, authors name, and at the foot of the page: the name of degree, Your University, and the year of presentation. Something like this:

<br>

# MSc. Thesis example VIBOT

<br>

Robert Martí

<br>

Department of Computer Architecture and Technology
University of Girona

<br>

A Thesis Submitted for the Degree of MSc Erasmus Mundus in Vision and Robotics (VIBOT)

· 2008 ·

### 1.2.1 References

You can reference other authors by using the *citecommand* [1]. You are encouraged to use bib files and let bibtex do the job for you.

# Chapter 2

# Background

## 2.1  Introduction

Many scientific subject areas (i.e. medicine, remote sensing) have to deal with the problem of image correspondence (also known as image registration or image matching). This problem arises when there is a need to extract information about an object by comparing a set of different images in which this object appears. Most of the time this comparison cannot be performed easily because of differences between images such as viewpoint changes, use of different sensors, images taken at different times, and changes in general imaging conditions. Given such differences, a way to align different images of the same object is needed. In medical applications, and specifically in mammography, radiologists have to deal with all the above problems. For example, a way to detect abnormal structures in the breast is to compare temporal (images of the same breast taken at different times) or contralateral (using left-right breast) mammograms.

As images are taken under different conditions (i.e. film sensitivity, radiation exposure, breast compression and patient movement) and/or time intervals (e.g. screening interval of three years) the structure of the breast is likely to suffer some changes, although an overall similarity will be maintained. A way to detect those changes is to align the images and compare the results using, for instance, simple image subtraction. In addition, as described in the previous chapter, different imaging modalities are being used to provide a better understanding of a region of interest. The first step to incorporate information from those modalities is to align them to be able to establish areas of correspondence.

An image correspondence method is based on finding a mapping function ($f(x, y)$) that maps each coordinate of one image ($A$) into another image ($B$).

$$B(x', y') = A(f(x, y)) \qquad (2.1)$$

which maps spatial coordinates $(x, y)$ in the reference image $A$ to coordinates $(x', y')$ in the warped (also referred to as target) image $B$. Usually the function $f$ is expressed as two (or three in three dimensions) separate functions $f_x$ and $f_y$ (and $f_z$). This review describes image registration in two dimensions. However extrapolation to 3D is often straightforward and is described in detail where appropriate.

A general methodology for image alignment typically follows those steps:

1. *Selection and extraction of features.* A registration method can be based on matching different image primitives. For instance points, edges, ridges, surfaces, or a whole image. The selection of a suitable primitive is a trade off between the information it provides and its complexity. For instance, raw pixels with only intensity information provide a straight-forward comparison. On the other hand, taking the whole image provides more (and more complex) information. Primitives will be described by a set of features extracted from them. That is the case, for instance, of shape features extracted from regions or curvature values extracted from linear structures. There is an extensive literature about feature extraction, the reader is referred to Chapter 4 where a review on that subject is given.

2. *Similarity metric.* A way to measure the similarity between the extracted features is needed. Many different similarity metrics have been proposed and their suitability depends on the chosen feature. The reader is referred to Chapter 5 where similarity measures are reviewed and a novel measure is proposed.

3. *Selection of the mapping function and estimation of its parameters.* The complexity of a mapping function should be determined depending on the application field and type of misalignment we are dealing with. Once a mapping function is selected its parameters need to be estimated. This requires a definition of an optimum search space and search strategy which is often related to optimisation of a cost function related to a similarity measure previously mentioned.

4. *Alignment of images using the mapping function.* When the function parameters are known we are able to transform an image in order to minimise the initial misalignment.

### 2.1.1 Interpolation

Usually transformed coordinates $(x', y')$ in the warped image do not match a given coordinate grid (i.e. they are non-integer values) and interpolation is needed. Different interpolation

methods have been used in image registration: nearest neighbour, tri-linear and partial volume distribution. The interpolation problem is graphically represented in Figure 2.1.

Figure 2.1: Interpolation of corresponding coordinates

Nearest neighbour interpolation assigns the intensity value of $m = (x', y')$ to its nearest point:

$$B(m) = B(\_\forall imin(d_E[m, n_i]))$$ (2.2)

where $d_E$ indicates Euclidean distance and $n_i$ are the nearest neighbours. The nearest neighbour approach provides good accuracy and it is easy to compute.

Trilinear interpolation assigns to $m$ the value of the weighted $(w_i)$ sum of the neighbouring pixels.

$$B(m) = \sum_i w_i B(n_i)$$ (2.3)

where each weight $w_i$ is related to the distance from $m$ (see Figure 2.1)

$$w_1 = (1 - dx)(1 - dy) \quad w_2 = (1 - dx)dy \quad w_3 = dx(1 - dy) \quad w_4 = dxdy$$ (2.4)

This interpolation creates a new intensity value, which could introduce undesirable effects in the intensity distribution. This fact is regarded as its main drawback, and can be solved using a partial volume distribution. This method adds the weight of each neighbour $(w_i$ values) at the intensity value in the histogram distribution, without creating an additional intensity value.

# Appendix A

# The first appendix

If you need to add any appendix, do it here... Etc.

# Bibliography

[1] Witold Pokorski and Graham G. Ross. Flat directions, string compactification and three generation models. 1998.