Spatio-Temporal Scale-Space Theory

DANIEL FAGERSTRÖM

Doctoral Thesis
Stockholm, Sweden 2011
Akademisk avhandling som med tillstånd av Kungl Tekniska högskolan framlägges till offentlig granskning för avläggande av teknologie doktorsexamen i datalogi fredagen den 10 juni 2011 klockan 10.00 i sal D3, Lindstedtsvägen 5, KTH, Stockholm.

© Daniel Fagerström, maj 2011

Tryck: E-print AB
Abstract

This thesis addresses two important topics in developing a systematic space-time geometric approach to real-time, low-level motion vision. The first one concerns measuring of image flow, while the second one focuses on how to find low level features.

We argue for studying motion vision in terms of space-time geometry rather than in terms of two (or a few) consecutive image frames. The use of Galilean Geometry and Galilean similarity geometry for this purpose is motivated and relevant geometrical background is reviewed.

In order to measure the visual signal in a way that respects the geometry of the situation and the causal nature of time, we argue that a time causal Galilean spatio-temporal scale-space is needed. The scale-space axioms are chosen so that they generalize popular axiomatizations of spatial scale-space to spatio-temporal geometries.

To be able to derive the scale-space, an infinitesimal framework for scale-spaces that respects a more general class of Lie groups (compared to previous theory) is developed and applied.

Perhaps surprisingly, we find that with the chosen axiomatization, a time causal Galilean scale-space is not possible as an evolution process on space and time. However, it is possible on space and memory. We argue that this actually is a more accurate and realistic model of motion vision.

While the derivation of the time causal Galilean spatio-temporal scale-spaces requires some exotic mathematics, the end result is as simple as one possibly could hope for and a natural extension of spatial scale-spaces. The unique infinitesimally generated scale-space is an ordinary diffusion equation with drift on memory and a diffusion equation on space. The drift is used for velocity adaption, the “velocity adaption” part of Galilean geometry (the Galilean boost) and the temporal scale-space acts as memory.

Lifting the restriction of infinitesimally generated scale spaces, we arrive at a new family of scale-spaces. These are generated by a family of fractional differential evolution equations that generalize the ordinary diffusion equation. The same type of evolution equations have recently become popular in research in e.g. financial and physical modeling.

The second major topic in this thesis is extraction of features from image flow. A set of low-level features can be derived by classifying basic Galilean differential invariants. We proceed to derive invariants for two main cases: when the spatio-temporal gradient cuts the image plane and when it is tangent to the image plane. The former case corresponds to isophote curve motion and the later to creation and disappearance of image structure, a case that is not well captured by the theory of optical flow.

The Galilean differential invariants that are derived are equivalent with curl, divergence, deformation and acceleration. These invariants are normally calculated in terms of optical flow, but here they are instead calculated directly from the the spatio-temporal image.
Acknowledgments

First of all I am grateful to my supervisor Jan-Olof Eklundh for his continuous support, all interesting discussions, for creating such a great research environment and for pushing me to at last finish my thesis. Secondly I want to thank Tony Lindeberg who inspired me to study scale-space theory and for our work together on the first article in this thesis. I also want to thank Ambjörm Naeve for introducing me to the wonderful world of geometry and Ove Franzén at Uppsala University for making me interested in vision research in the first place.

Thanks also to: Lars Bretznér, Mattias Lindström, Pär Fornland, Peter Nillius, Mårten Björkman, Peter Nordlund, Anders Orebäck, Tomas Uhlin, Harald Winroth, Stefan Carlsson, Fredrik Bergholm and all other people, past and present in the group for all interesting discussions and for making CVAP such a stimulating environment.

Finally I would like to express my gratitude for the support and encouragement I have received from my family.
# Contents

## 1 Introduction
- 1.1 Optical Flow .................................................. 1
- 1.2 Local Spatio-Temporal Image Structure ....................... 2
- 1.3 Overview of Our Approach .................................. 3
- 1.4 Articles ....................................................... 5
- 1.5 Contributions .................................................. 5

## 2 Geometry of Visual Information
- 2.1 Introduction .................................................... 7
- 2.2 Visual Input ................................................... 7
- 2.3 Visual Transformations ........................................ 10
- 2.4 Spatio-Temporal Geometry .................................... 17
- 2.5 Invariance and Covariance ................................... 20
- 2.6 Infinitesimal Invariance ...................................... 21
- 2.7 Generalizations ................................................. 24

## 3 Geometrical Measurement
- 3.1 Introduction .................................................... 27
- 3.2 Extended Point .................................................. 28
- 3.3 Covariance ....................................................... 31
- 3.4 Cascade Property ............................................... 32
- 3.5 Infinitesimal Generators ....................................... 34
- 3.6 Positivity and Gray Level Invariance ......................... 38
- 3.7 Infinitesimal C Scale-Space .................................. 39

## 4 Spatio-Temporal Scale-Spaces
- 4.1 Introduction .................................................... 41
- 4.2 The Affine Line .................................................. 42
- 4.3 Similarity ....................................................... 43
- 4.4 Galilean Similarity ............................................. 44
- 4.5 Time Causal Scale-Spaces ..................................... 45
4.6 Time Causal Galilean Scale-Spaces ........................................ 47
4.7 Discussion ................................................................. 48

5 Article Summaries .......................................................... 51
  5.1 Article I: Scale-Space with Causal Time Direction ...................... 51
  5.2 Article II: Temporal Scale Spaces ....................................... 54
  5.3 Article III: Spatio-Temporal Scale-Spaces ............................... 55
  5.4 Article IV: Galilean Differential Geometry of Moving Images ....... 57

6 Conclusion and Open Issues .............................................. 61
  6.1 Conclusion ............................................................... 61
  6.2 Open Issues ............................................................. 62

A Categories ................................................................. 65

B Calculus on Manifolds ...................................................... 67
  B.1 Topological Prerequisites ................................................ 67
  B.2 Manifolds ................................................................. 68
  B.3 Tangent Bundle ........................................................... 69
  B.4 Bundles ................................................................. 70
  B.5 Embeddings and Immersions ............................................. 71
  B.6 Foliations ............................................................... 71

C Distributions ............................................................... 73

D Lie Groups ................................................................. 77
  D.1 Abstract Groups ........................................................ 77
  D.2 Lie Groups .............................................................. 78
  D.3 Lie Algebra .......................................................... 78

Bibliography ................................................................. 81
Chapter 1

Introduction

Motion is the most important cue in biological vision. There are animals that lack stereopsis or color vision or both, but no seeing animal is without visual motion sensitivity (Nakayama 1985). Why is that? Vision as a distal sense is inherently connected to motion, it would be of no evolutionary value to be able to detect a distant object without the ability to approach or avoid it. Some animals, e.g. frogs, seem to be dependent on visual motion to the degree that they never attend to non-moving objects (Arbib 1987). Coherent motion, or common fate in gestalt terminology (Koffka 1935), distinguish object from background with far less assumptions than any static image properties, like edges, color or texture. The perceptual strength of common fate can vividly be seen for camouflaged animals in their natural environment: while standing still they are nearly impossible to see but as soon as they start to move we immediately attend to them. The effect is so strong that we even can detect partly occluded motion, e.g. while the camouflaged animal is partly hidden behind foliage.

1.1 Optical Flow

The dominating approach to computational visual motion processing (reviewed in Barron, Fleet & Beauchemin (1994) and Mitiche & Bouthemy (1996)) is to first compute the optical flow vector field, i.e. the velocity vectors of the particles in the visual observer’s field of view, projected on its visual sensor area. Then various properties of the surrounding scene can be computed from the optical flow field. Ego-motion can, under certain circumstances, be computed from the global shape of the field, object boundaries from discontinuities in the field, and surface shape and motion for rigid objects, can be computed from the local differential structure of the field (Koenderink & van Doorn 1975, Koenderink & van Doorn 1976). The main ideas in this approach were introduced by Gibson (1955).

Unfortunately the computation of the optical flow field leads to a number of well known difficulties. The input is the projected (gray-level) image of the surroundings
as a function of time, i.e. a three-dimensional structure, that henceforth will be denoted a movie. It is in general not possible to uniquely identify what path through the movie is a projection of a certain object point. Thus, further assumptions are needed, the most common one is the brightness constancy assumption (Horn & Schunck 1981), that the projection of each object point has a constant gray level. The brightness constancy assumption breaks down if the light changes, if the object have non-Lambertian reflection, or, if it has specular reflections (Verri & Poggio 1989).

However, the problem is still under-determined, generically. Except at local extrema in the gray-level image, points with a certain gray-level lie along curves, and these curves sweep out surfaces in the movie. A point along such a curve can therefore correspond to any point on the surface at later instants of time. This is a slightly more general formulation of the so called aperture problem (Wallach 1935, Marr 1982). The aperture problem is usually treated by invoking additional constraints e.g. regularization assumptions, such as smoothly varying brightness patterns, or parametrized surface models and trajectory models, leading to least-square methods applied in small image regions. Beside the questionable validity of these assumptions they lead to inferior results near motion boundaries, i.e. the regions that carry most information about object boundaries. The behavior when new image structure appears or old structure disappears is also undefined.

There are additional difficulties. Hence, it might also be more appropriate to have a multi-valued description of local image velocity to be able to describe transparent motion, e.g. motion behind smoke, behind a foliage or behind a fence. Not surprisingly, state of the art algorithms for computing optical flow are fairly complicated.

To conclude: the relation between object motion and image motion is complicated and is in general not possible to determine from local information. Optical flow based models impose more structure on the movie than is actually there and it cannot describe some of structure that actually is there: appearance, disappearance and multi-valued motion.

1.2 Local Spatio-Temporal Image Structure

My conclusion from the various shortcomings of the optical flow approach is that optical flow is too much of a high level feature to be appropriate as a basic building block for motion analysis.

An alternative approach for visual motion analysis is to directly analyze the geometrical structure of the spatio-temporal input image, thereby avoiding the detour through the optic flow estimation step (Yamamoto 1988, Bolles, Baker & Marimont 1987, Zetzsche & Barth 1991, Jahne 1993).

A systematic study of the local image structure, in the context of scale-space theory, has been pursued by Florack (Florack, ter Haar Romeny, Koenderink & Viergever 1993, Florack, ter Haar Romeny, Koenderink & Viergever 1994).
1.3. OVERVIEW OF OUR APPROACH

basic idea is to find all descriptors of differential image structure that are invariant to rotation and translation (the Euclidean group). The choice of Euclidean invariance reflects that the image structures should be possible to recognize in spite of (small) camera translations and rotations around the optical axis. This theory embeds many of the operators previously used in computer vision, such as Canny’s edge detector, Laplacian zero-crossings, blobs, isophote curvature and as well enabling the discovery of new ones.

Our aim is to carry through a program similar to Florack (Florack et al. 1993, Florack et al. 1994), for moving images, and thus get a systematic theory about local spatio-temporal visual structure and how to measure it.

Our work is also inspired by results from neuro physiology. From the spatio-temporal shape of the receptive fields in LGN and V1 (DeAngelis, Ohzawa & Freeman 1995, DeAngelis, Ghose, Ohzawa & Freeman 1999) it looks like the visual system might calculate mixed spatio-temporal derivatives from the visual input.

Experimental results from gestalt theory about apparent motion are hard to explain in a two-frame optical flow setting but makes perfect sense with motion oriented spatio-temporal filters.

A family of velocity adapted spatio-temporal filters can explain transparent layered motion in a better way than a single valued optical flow sensor.

1.3 Overview of Our Approach

We will carry through this program in a number of steps. First, in Chapter 2, we discuss the structure of the visual input, with a focus on those aspects that we believe are most important for low-level motion vision.

There are two main perspectives for modeling the visual input. One is the object centered perspective that starts from a model of the objects and some of their properties i.e. motion or surface reflectance, and then describes how these properties are projected onto the retina. This approach is common in reconstructive computer vision, where the main goal is to find an inverse to the image formation projection and thus be able to reconstruct a model of the objects from the image data. Another approach is an image centered perspective where one starts by postulating fairly general properties of the image, such as smoothness, linearity, diverse symmetry properties, or maybe certain statistical properties. This is the approach that we will adapt, and it has been more popular in bottom-up oriented approaches to understand vision.

A central question is: what kind of structure are we interested in? There is of course no unique answer, the answer depends on what we need to know about the environment. In this work, we are interested in mobile observers, for which properties of the possibly moving objects in their surroundings are of fundamental importance.

Euclidean invariance in the image plane is as useful in dynamic, as in static imagery. Image properties should not be dependent on when we choose to measure
them (invariance under time translations). In space-time there are several symmetry groups that might be fruitful to consider. The local average velocity contains only information about the ego motion while information about the three-dimensional structure of the environment is contained in the local change in the motion field (Koenderink & van Doorn 1975). It is therefore natural to separate these two aspects of the visual input and try to find properties that are invariant to local average velocity (Galilean boost). There are other properties in the space-time image that only depends on the ego motion (Koenderink & van Doorn 1981), but this is at least a first step. We thus search for properties that are invariant to the 2+1 dimensional Galilean group. We review the necessary background in group theory and geometry and give explicit mathematical descriptions of the input, in terms of geometric function-spaces, and thus formalize the input models that we will use in the sequel.

If the visually guided system wants to make any use of the visual input, it has first to measure it. In Chapter 3 we formulate a set of requirements for measurement in geometrical spaces. We follow Florack’s (Florack 1997) distribution theory based approach. The resulting framework can be considered as a slight generalization of existing linear scale-space theories. A measurement theory consists of an input space, a space of measurement devices and a way of applying these to each other. As our main interest lies in local properties we want measurement devices to be as point-like as possible, but to be physically realizable. The measurements must be done over a non-vanishing volume of the input space. For a linear measurement theory, this means that the measurement devices can be modeled by non negative kernels that are localized in some appropriate way. Each transformation on the input space corresponds to a dual transformation on a measurement kernel and the measurement kernel typically changes shape when it is transformed. We don’t want to favor any particular transformation, thus an important sub-problem is to find a family of measurement kernels that is as compatible with the structure of the input as possible. The main result of this study is a set of infinitesimal constraints for the measurement kernels for a given input geometry.

The framework is then applied to a couple of spatial and spatio-temporal geometries in Chapter 4. First all one-dimensional scale-spaces on the affine line are derived. Then spatio-temporal scale-spaces with a focus on Galilean scale-space with causality in the temporal domain.

The chapters referred to above gives a background to (and for spatio-temporal scale-space extends) the included articles that are summarized in Chapter 5.

We conclude, in Chapter 6, by summarizing our results, discussing possible further work, speculating about the hereby presented framework relationship to biological vision, and by discussing some of the remaining issues and some possible generalizations.
1.4 Articles

In article I (Lindeberg & Fagerström 1996), time causal temporal scale-spaces are developed from the scale-space axioms from (Lindeberg 1990, Lindeberg 1993). The resulting scale-spaces are either discrete in time or scale.

Article II (Fagerström 2003, Fagerström 2005) is also about time causal temporal scale-spaces. Here the set of axioms is modified to enable scale-spaces that are continuous in both the temporal and the scale direction. Like previous approaches, (Koenderink 1988, Lindeberg & Fagerström 1996, Salden, ter Haar Romeny & Viergever 1998), we postulate temporal causality: the system can not access future input. A property of temporal measurement that has received less attention, is the fact that a measurement system can not access past data either. It can only access memories of past measurements, i.e. a theory about temporal measurement must also include some kind of rudimentary memory. We discuss the consequences of these requirements and show that there is a family of temporal scale-spaces that uses the temporal scale-dimension as memory of past measurements. We conclude the paper by discussing numerical issues.

In article III (Fagerström 2007) the theory is extended to time causal velocity adapted spatio-temporal scale-spaces. To make this possible an infinitesimal framework for scale-spaces with respect to different Lie geometries is developed. This framework is applied to a number of input geometries: 1-dimensional affine geometry, Euclidean similarity geometry and Galilean similarity geometry, leading to the corresponding scale spaces. Especially interesting in the context of motion processing is the measurement of temporal data.

In the last article IV (Fagerström 2004), we study two different classes of local invariants for functions over a Galilean space, differential invariants, i.e. functions over derivatives in a point, that are invariant w.r.t. Galilean transformations, and level set invariants, that are invariant over (global) monotone intensity transformations as well.

1.5 Contributions

For Article I, I took the initiative to study causal temporal scale-spaces and developed most of the results for scale-spaces based on cascaded truncated exponential and moving average filters (in Section 3-5). Tony Lindeberg developed the axiomatization for temporal scale-spaces used in the article, the Poisson kernel based scale-space, led the work and wrote the article.

For the rest of the articles I’ve developed the theory and is responsible for all results.
Chapter 2

Geometry of Visual Information

2.1 Introduction

What is useful to see? The amount of possible visual information is so vast that a “general” theory about vision is currently far beyond reach and it seems questionable if such a theory is feasible at all. In this work we will focus on an ecologically based approach: there is an animal (or possibly a robot) that sees an environment. As the measurement and computation of visual structure takes “hardware” and energy resources, one can assume that evolution favors development of visual competences that is useful for the behavior of the animal. A similar argument from an economical perspective applies to robots. The main developer of the ecological approach to visual perception was Gibson (1979). Ecological arguments are also used in computer vision research, especially within the sub-field of active vision (see e.g. Ballard (1991) and Pahlavan, Uhlin & Eklundh (1993)). In this chapter we will first discuss the structure of some of the basic parts of the visual information that can be useful for animals with human-like vision. The main concepts are from Gibson (1979) but we will make the mathematical structure more explicit. We will then focus our work on finding a mathematical description of those aspects of the visual structure that we believe are important for understanding early spatio-temporal vision. Our approach is geometry based and the goal is to describe the input in terms of an appropriate geometric space.

2.2 Visual Input

The environment consists of a medium, air for humans, and water for water living creatures and substance, that is solid matter. For us water is somewhere in between medium and substance, but more like substance. The medium is characterized of that one can move through it, that it transmits light, sound, and odor, it is also quite homogeneous - without structure. Substance does (in most cases) not permit easy motion for animals or transmission of light, sound or odor. Substance also
tends to be fairly persistent, most of the time its configuration changes just a little from one moment of time to the next. The environment lies in $3 + 1$-dimensional space-time $\mathbb{R}^3 \times \mathbb{R}$.

The main source for visual perception is the interface between medium and substance: surfaces, which are 2-dimensional continuous manifolds, $M^2$, embedded in space. Surfaces have a layout that consists of overall shape, texture and reflectance properties. Each substance has characteristic reflectance properties. The reflectance at a given point $x \in M^2$ at the surface can be described by a bidirectional reflectance function (BRDF), $B : M^2 \times S^2 \times S^2 \times \mathbb{R} \to \mathbb{R}$, $(x, \theta_i, \theta_e, \lambda) \mapsto B(x, \theta_i, \theta_e, \lambda)$, given a point light source with the wavelength $\lambda$ in the direction $\theta_i \in S^2$. $S^2$ is the surface of a sphere, a 2-dimensional manifold\(^1\) of directions, (see e.g. Horn (1986)). The BRDF describes the ratio between the amount of incoming light and the reflected brightness seen from the direction $\theta_e \in S^2$. We will only discuss grey value vision in the sequel, so we drop the wavelength parameter $\lambda$.

For most naturally occurring surfaces the BRDF is smooth and rotationally invariant. It can often be assumed that the BRDF is approximately constant, for small changes in viewing direction. This assumption holds for most viewing directions for a large range of naturally occurring surfaces, but it breaks down near viewing directions that give rise to specularities.

As most substance is fairly persistent their surfaces are persistent as well. Light sources emit light of about the same intensity in a wide range of directions. If we follow a ray of light from its source it interacts with the atmosphere (medium) and a fraction of its intensity is scattered away in other directions. If it hits a surface, part of its intensity is absorbed by the substance and the rest is reemitted from the surface and the intensity in the different directions depends on the surfaces reflectance properties. The reemitted light might in turn interact with other surfaces. The end result is that each point in the medium has incoming light from all directions, Gibson calls this incoming light ambient light. The ambient light is a function over a two-dimensional manifold of directions, $S^2 \to \mathbb{R}$, $x \mapsto u(x)$. There is ambient light in each point in the medium, and the totality of ambient light for all spatio-temporal points is called the ambient optical array, $\mathbb{R} \times \mathbb{R}^3 \to (S^2 \to \mathbb{R})$, $(t, X) \mapsto u_{t,X} : S^2 \to \mathbb{R}$, this is, (except for the color), all that could possibly be seen.

In the rest of this chapter we will try to make the mathematical structure of the ambient optical array more explicit. We do this by separating the problem in two parts: describing the structure of the ambient light at a certain point, and describing how the ambient light changes from one spatial point to another and from one moment in time to the next.\(^2\)

\(^1\)A manifold is a space that locally is like a Euclidean space, i.e. curves and surfaces, see Appendix B.2.

\(^2\)This is not the space-time geometry approach that I promised in the introduction, but I think it is easier to start from a more traditional perspective before we introduce space-time geometry.
2.2. VISUAL INPUT

Linearity of the Signal

The ambient light can be considered as a linear space of functions. Consider a room with two spotlights. Turn on one of them, then the intensity of the ambient light at a certain point is \( u : M^2 \rightarrow \mathbb{R} \). If we change the light intensity from the spotlight with a factor \( c \in \mathbb{R} \), the ambient light will be \( cu \), where \((cu)(x) = c(u(x))\). Turn off the spotlight and turn on the other one, then the ambient light at the point will be \( v : M^2 \rightarrow \mathbb{R} \). If both spotlights are turned on at once we get \( u + v \), where \((u + v)(x) = u(x) + v(x)\).

Structure of the Ambient Light

The ambient light of a point is a mosaic of the projections of the reflected light of the nearest facing surfaces in the environment. The borders of the patches in the mosaic are projections of the contours of the objects in the surroundings. A contour of an object with respect to a viewpoint are all points at its surface that are tangential to the viewpoint. There are two forms of contours: boundary contours that separates surfaces from different objects and self occlusion contours.

There are thus two different categories of points on the ambient light manifold: interior points that have a neighborhood such that all points in the neighborhood originate from one surface, and contour points that lie along curves. Border contour points have points from more than one surface in its immediate neighborhood. Self occlusion contour points have points from two distinct neighborhoods in its immediate neighborhood.

Local Descriptors

In this thesis we will only consider local properties, i.e. properties that only need information from the immediate neighborhood around a point. These local properties describe the structure around an internal point or the structure around a contour point. There are of course an abundance of non-local properties that can be useful for a seeing system, but as these typically will get input from several different surfaces in the environment, their correlation structure is immensely much more complicated. Direct measurement of multi-point properties also leads to a combinatorial explosion of the number of possible measurements and seems to be computationally infeasible to handle for a biological or artificial vision system. We believe, from complexity considerations, that all but the most specialized vision systems, must measure non-local properties from refined local properties rather than from the raw input.

There is no reason to believe that the ambient light is a smooth function, but as non-smooth functions are so much more complicated to analyze, we will define measurement is such a way that we can continue our discussion as if the ambient light really is smooth\(^3\). We will actually, in Chapter 3, define the visual input to be a distribution, which allow
CHAPTER 2. GEOMETRY OF VISUAL INFORMATION

Eye Model

A visually guided animal living in an environment needs some way to extract information from the ambient optical array. An idealized way of doing this is to use a pinhole camera. A pinhole camera is a box or some other kind of volumetric container with a point sized hole on its surface. Thus part of the ambient light at this point is projected on the surface facing the hole at the inside of the pinhole camera, we call this projection a picture. The projection can be on a curved surface as in most animal eyes or on a planar surface as in a camera. In both cases the surface is a two-dimensional manifold. The camera also has a position and a rotation at each moment in time so the picture can be described as a function $\mathbb{R} \times \mathbb{R}^3 \times \text{SO}(3) \rightarrow (M^2 \rightarrow \mathbb{R})$, $(t, X, R) \mapsto u(t, X, R) : M^2 \rightarrow \mathbb{R}$. Most of the time we will drop the explicit dependency on position, time and rotation and just write $M^2 \rightarrow \mathbb{R}$, $x \mapsto u(x)$. The pictures development over time is called a movie $M^2 \times \mathbb{R} \rightarrow \mathbb{R}$, $(t, x) \mapsto u(t, x)$. As we are mainly interested in local aspects and consider the ambient light being a linear space of functions our basic object of study will be $L(U \rightarrow \mathbb{R})$ where $U \subset \mathbb{R} \times M$ is an open set.

2.3 Visual Transformations

This far we have decided that we are only interested in local properties of the ambient light. We have also noted that there are two main categories of points in the ambient light manifold: interior points and contour points. Now we will study how the ambient light changes from one spatio-temporal point to a nearby point, and thus learn about the local structure of the ambient optical array. Here we also stick to local structures as these are much easier to handle. We will see what the environment looks like from differently placed cameras in a static environment, what an object looks like from different viewpoints. We will also discuss what happens in a dynamical environment, when both the camera and the objects in the environment move around and when the light changes.

One of the more important tasks for the visual system, is to recognize different classes of objects and events from visual input. If we look at an object, the visual input from it will change if we change the viewing position and if the light changes, but in most cases we will still be able to recognize the object. Some kind of information, that is derivable from the visual input must evidently stay unchanged, such aspects are called invariants. We can recognize a persons face even if he or she changes facial expression and we can recognize a certain facial expression on several people. We can recognize walking, running, swimming, flying and so on irrespective of who performs it. In all these examples there is an invariant aspect and a variant aspect of the visual input.

---

For as non-smooth input that one could possibly want. These extra complications seem however unnecessary for our current discussion.
2.3. VISUAL TRANSFORMATIONS

Spatial Change

Consider a point in the environment that is projected as an interior point for a certain camera. If we move the camera just a little the point will still be projected as an interior point. Furthermore all points in a small enough neighborhood around the point will still be visible. The change can thus be described as a mapping from a neighborhood in the picture to another neighborhood, and this mapping will describe how the projection of a point in the environment as well as its intensity value, will change. The mapping will also be invertible as the underlying physical process is reversible, we could have started with the camera in the second position and moved the camera back to the first position. This kind of invertible mappings on a space are called transformations, we need some basic facts about transformations on function spaces to be able to continue our discussion.

Transformations

An invertible map that preserve the structure of a space is called a automorphism (see Appendix A), and the space of automorphisms on the space $X$ is denoted $\text{Aut}(X)$. We are mainly interested in spaces that are manifolds. An automorphism on a manifold is called a diffeomorphism, and a diffeomorphism is a smooth map on a manifold with a smooth inverse.

Definition 2.3.1. The graph of the function $f: X \rightarrow Y$ is defined as

$$\Gamma_f = \{(x, y) \mid y = f(x), x \in X\}.$$ 

The space of graphs over the function space $X \rightarrow Y$ is denoted $\Gamma(X \rightarrow Y)$.

Graphs in $\Gamma(C^\infty(X,Y))$ are $\dim(X)$ dimensional smooth submanifolds of the space $X \times Y$. Observe that there are submanifolds that are not graphs to any functions.

We continue by defining a couple of classes of transformations on function spaces, (see Olver (1995) for details).

Definition 2.3.2. Let $F: \text{Aut}(X \times Y)$ such that

$$(x, y) \mapsto F(x, y) = (F_b(x, y), F_f(x, y)),$$

if $F$ furthermore is a automorphism on graphs, i.e. $\forall f: X \rightarrow Y, \exists g: X \rightarrow Y$ such that $F(\Gamma_f) = \Gamma_g$, the $f$ induces a automorphism, satisfying $F(f) = g$ denoted a point transformation on the corresponding function space $F: \text{Aut}(X \rightarrow Y)$.

In our applications this is unnecessarily general. The part of the transformation that acts on the domain of the function does not need to be dependent of the value of the function.
CHAPTER 2. GEOMETRY OF VISUAL INFORMATION

Definition 2.3.3. A fiber preserving transformation is a point transformation induced by a mapping $F : \text{Aut}(X \times Y)$, such that

$$(x, y) \mapsto F(x, y) = (F_b(x), F_f(x, y)).$$

Note that $F_b$ and $y \mapsto F_f(x, y)$ must be automorphisms from the definition and therefore the inverse must be fiber preserving.

The notion “fiber preserving transformation” comes from that the graph of a function can be considered as a section of a fiber bundle (see Appendix B.4). A (smooth) fiber bundle is a construction where one puts some kind of structure, called the fiber, $F$, on each point of a manifold, called the base space, $B$, in such a way that the total space, $E$, also becomes a manifold. There is a projection, $\pi : E \to B$, from the total space to the base space. For each neighborhood, $U \subset B$, of the base space, the bundle, $\pi^{-1}(U) \subset E$ is isomorphic to the Cartesian product of the neighborhood and the fiber, $U \times F$. A section of a smooth bundle is a smooth map, $\sigma : B \to E$, that chooses a value in the fiber for each point on the base space, $\pi \circ \sigma = id_B$. The space of smooth sections over the fiber bundle $E \to B$, is denoted $C^\infty(E \to B)$ or $C^\infty(E)$. The graph of a map $f : X \to Y$, can thus be identified with a section of the smooth bundle $\pi : X \times Y \to X$. A fiber preserving transformation never mixes the fibers as the point transformations can do.

Definition 2.3.4. A base transformation is a point transformation induced by $F : \text{Aut}(X \times Y)$, $(x, y) \mapsto F(x, y) = (F_b(x), y)$.

Most of the transformations we will use are base transformations. Transformations that are defined for the domain of a function space induces obviously an unique base transformation on the function space: $\text{Aut}(X) \to \text{Aut}(X \times Y)$, $F_b \mapsto (F_b, id)$, with a slight abuse of notation we will use the same symbol for both transformations.

Each of the sets: $\text{Aut}(X)$, $\text{Diff}(X)$, point-, fiber preserving- and base transformations are groups under composition of maps (see Appendix D for definitions).

Example 2.3.1. Let $F, G \in \text{Aut}(X \times Y)$ be fiber preserving transformations, then their composition

$$F \circ G(x, y) = (F_b \circ G_b(x), F_f(G_b(x), G_f(x, y))),$$

is also fiber preserving. Therefore fiber preserving transformations form a group as they are closed under composition and inversion. The fiber preserving transformations are automorphisms on fiber bundles.

Equipped with these definitions we can be more precise about the form of the transformations around internal points, discussed above. The transformations are also smooth in both directions (as long as the camera motion and the environment around the interior point is small enough), so the transformations are diffeomorphisms on $U \to \mathbb{R}$ where $U \subset M^2$. They should also be fiber preserving or maybe base transformations. Along contours the situation is more complicated. When
the camera moves some points may become hidden and new points may appear. A third class of events are singularities - when new contours appear or old disappear. We will see that a spatio-temporal perspective simplifies the handling of contours.

**Lie Transformation Groups**

The group of fiber preserving diffeomorphisms over some function space is infinite dimensional and has too little structure for our needs. We will use Lie transformation groups instead. A Lie group is a group that also have manifold structure (see Appendix D for more details), and where the group operation and the inverse, are smooth maps. The manifold structure of Lie groups means that there is a Euclidean parametrization of the group in a neighborhood of each point, and also that there is a local linearization of the group around each point. These properties simplify the discussion considerably.

**Examples 2.3.2.** We exemplify with a number of Lie groups that we will use henceforth.

1. The set of $n$-dimensional vectors, $\mathbb{R}^n$, under vector addition, form a group.
2. The set of non-negative $n$-dimensional vectors, $\mathbb{R}_+^n$, under point wise multiplication, form a group.
3. The set of non-singular $n \times n$ matrices under matrix multiplication, form a group: the general linear group, $\text{GL}(n)$.
4. $n \times n$ matrices with determinant 1 form a compact subgroup of $\text{GL}(n)$ called the special linear group $\text{SL}(n)$.
5. The $n$-dimensional orthogonal group, $O(n)$ is the subgroup of $\text{GL}(n)$, such that the elements $A \in O(n)$ satisfies $AA^t = I$. $\text{SO}(n)$ is the subgroup of $O(n)$ with determinant 1, especially can elements in $\text{SO}(2)$, two dimensional rotations, be represented by matrices

$$R(\theta) = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}, \quad \theta \in \mathbb{R},$$

and $R(\theta)R(\gamma) = R(\theta + \gamma)$.

**Definitions 2.3.5.** If a Lie group $G$ is acting on a smooth manifold $M$ by means of a Lie group homomorphism $\sigma : G \to \text{Diff}(M)$, the set of transformations with map composition as group multiplication \{\sigma(G), \circ\} is called a Lie transformation group. If the choice of homomorphism is evident from the context, we use the notation $gx$ or $g \cdot x$, where $g \in G$ and $x \in M$. The triple $(M, G, \sigma)$ is called a differentiable $G$-space. A smooth mapping $f : M \to N$ is called a differential $G$-space morphism, if there exists a homomorphism $\phi : G \to G'$, such that $f(gx) = \phi(g)f(x)$, for all $g \in G, x \in M$. A $G$-space isomorphism is an invertible $G$-space morphism. A more
restricted form of morphism, when \( \phi = id \), and \( G \) thus acts on both \( M \) and \( N \), \( f(gx) = gf(x) \), is called equivariance.

**Example 2.3.3.** Any Lie group \( G \) can act on itself by mean of left translation, right translation and inner automorphism, \( l, r, a : G \rightarrow \text{Diff}(G) \),

\[
l_g(x) = gx, \quad r_g(x) = xg, \quad a_g(x) = gxg^{-1}.
\]

An inner automorphism can be written as: \( a_g = l_g \circ r_g = r_g \circ l_g \). The inner auto-morphism can easily be seen to be a group-isomorphism with the inverse, \( (a_g)^{-1} = a_{g^{-1}} \).

**Definition 2.3.6.** Two sub-groups \( A, B \) of a group \( G \) are said to be conjugate if there exists \( g \in G \) such that \( a_g(A) = B \).

Conjugacy is an equivalence relation on groups.

**Examples 2.3.4.**

1. All the subgroups of \( \text{GL}(n) \) described above acts on \( \mathbb{R}^n \), with the identity homomorphism: \( \sigma(A)x = Ax \), where \( A \in \text{GL}(n) \), and \( x \in \mathbb{R}^n \).

2. The \( n \)-dimensional general affine group \( \text{GA}(n) \) acts on elements in the vector space \( x \in \mathbb{R}^n \) by \( Ax + v \), where \( A \in \text{GL}(n) \) and \( v \in \mathbb{R}^n \). The group operation is:

\[
(A_1, v_1) \cdot (A_2, v_2) = (A_1, v_1)(A_2x + v_2) = A_1A_2x + A_1v_2 + v_1 = (A_1A_2, v_1 + A_1v_2). 
\]

3. The \( n \)-dimensional Euclidean similarity group \( \text{ES}(n) \) is a subgroup of \( \text{GA}(n) \) where an element has the form: \( (\lambda R, v) \), \( \lambda \in \mathbb{R}_+ \), \( R \in \text{SO}(n) \) and \( v \in \mathbb{R}^n \). Elements of the form \( (\lambda I, 0) \), \( (R, 0) \) and \( (I, v) \) form the subgroups of scaling, rotation and translation, respectively.

4. The \( n \)-dimensional Euclidean group \( E(n) \) is a subgroup of \( \text{ES}(n) \) without scaling, \( \lambda = 1 \), an element has the form: \( (R, v) \), \( R \in \text{SO}(n) \) and \( v \in \mathbb{R}^n \).

The Euclidean similarity group and the general affine group have both central roles in vision science, and are important for our further studies of the spatial aspect of early vision. Consider a pinhole camera where the picture is projected on a planar surface, then two pictures taken from two different positions of a planar surface in the environment, are related by a projective transformation. If the distance to the surface is large compared to the size of the surface, the relations between pictures from two different camera positions can be approximated by an affine transformation. If the camera is translated and rotated within its picture plane, two pictures are related by a Euclidean transformation. In the mentioned examples we assumed that the BRDF was constant for each point at the surface.
2.3. VISUAL TRANSFORMATIONS

Structure of Groups

The structure of a group can often be described by decomposing the group as products or semi-direct products of its sub-groups.

Definitions 2.3.7. Two groups $G, H$ can be combined to a product group $G \times H$ equipped with a group operation defined by:

$$(g_1, h_1) \cdot (g_2, h_2) = (g_1 \cdot g_2, h_1 \cdot h_2),$$

$g_1, g_2 \in G, h_1, h_2 \in H$. A semi-direct product of two groups $G \rtimes_b H$ where $H$ acts on $G$ by means of a homomorphism $b : H \rightarrow \text{Aut}(G)$ is equipped with a group operation defined by:

$$(g_1, h_1) \cdot (g_2, h_2) = (g_1 \cdot b(h_1)(g_2), h_1 \cdot h_2).$$

Elements of the form $(g, \text{id})$ (and $(\text{id}, h)$) are subgroups of $G \rtimes H$ and isomorphic to $G$ (respectively $H$). $G$ is normal in $G \rtimes H$ and $hgh^{-1} = b(h)g$. This means that $(G \rtimes H)/G$ is a quotient group that is isomorphic to $H$.

Examples 2.3.5. $\text{GA}(n) \simeq \mathbb{R}^n \rtimes \text{GL}(n)$, where $e$ is the identity homomorphism $e(R)x = Rx$. We have also $\text{E}(n) \simeq \mathbb{R}^n \rtimes \text{SO}(n)$ and $\text{ES}(n) \simeq (\mathbb{R}^n \rtimes \mathbb{R}_+) \rtimes \text{SO}(n)$ where $b : \mathbb{R}_+ \rightarrow \text{Aut}(\mathbb{R}^n)$ and $b(\lambda)(v) = \lambda v$.

Lie Algebra

From a Lie group $G$ we get a corresponding infinitesimal object, its Lie algebra,

$$\mathcal{L}G = \mathfrak{g}.$$ 

The Lie algebra is the tangent space of $G$, $TG_e$ around the identity element $e$, together with an anti-commutative bilinear operator

$$\mathfrak{g} \times \mathfrak{g} \ni (v, w) \mapsto [v, w] \in \mathfrak{g},$$

the Lie bracket, that satisfy the Jacobi identity,

$$[[u, v], w] + [[v, w], u] + [[w, u], v] = 0$$

for $u, v, w \in \mathfrak{g}$, see e.g. (Onishchik 1993) for details. The tangent algebra of $G$ can be generated by differentiating one parameter subgroups $\mathbb{R} \ni t \mapsto g(t) \in G$ at the identity element $\partial tg(e)$. For Lie algebras of linear operators $A, B$, the Lie bracket is defined as

$$[A, B] = AB - BA.$$
CHAPTER 2. GEOMETRY OF VISUAL INFORMATION

From the Lie algebra the connected component around identity in the Lie group can be recreated with the exponential map \( \exp : \mathfrak{g} \to G \), for matrix Lie algebras the exponential map becomes the matrix exponential. For a Lie group homomorphism \( f : G_1 \to G_2 \) there is a corresponding Lie algebra homomorphism \( df : \mathfrak{g}_1 \to \mathfrak{g}_2 \) such that for \( v, w \in \mathfrak{g}_1 \),

\[
\exp([v, w]) = [\exp(v), \exp(w)].
\]

More generally, a smooth map between two manifolds \( f : M \to N \), induces a linear map

\[
M \ni x \mapsto df(x) : TM_x \to TN_{f(x)},
\]

between the tangent spaces of the manifolds, called the differential.

Space Geometry Lie Algebras

Now we continue by listing Lie algebras for the spatial geometries that we will use:

A base for the Lie algebra of translations is,

\[
t(n) = \{\partial_1, \ldots, \partial_n\},
\]

and all commutators are zero.

The affine line has the infinitesimal generators,

\[
\mathfrak{gl}(1) = t(1) \cup \{x\partial_x\},
\]

and the commutator

\[
[\partial_x, x\partial_x] = \partial_x.
\]

The Euclidean similarity group on \( \mathbb{R}^2 \) consists of translation in the plane rotation and scaling. Scaling is generated by

\[
s(2) = \{s = x_1\partial_1 + x_2\partial_2\}
\]

and rotation by

\[
\mathfrak{so}(2) = \{r = x_2\partial_1 - x_1\partial_2\},
\]

and the Euclidean similarity algebra by,

\[
\mathfrak{es}(2) = t(2) \cup s(2) \cup \mathfrak{so}(2),
\]

where the non-zero commutators are

\[
[\partial_j, s] = \partial_j, \quad [\partial_1, r] = -\partial_2 \quad \text{and} \quad [\partial_2, r] = \partial_1.
\]
2.4 Spatio-Temporal Geometry

When we add the temporal dimension, new sources of change appear: The objects in the environment move around, light sources change in intensity and position, the camera moves.

A common (sub) goal for research in visual motion is to find a mechanism that makes it possible to follow the image of points at a surface in the environment from one moment of time to the next. This means that an invertible map is supposed to describe the change over time. As we saw above this assumption only holds for environment around interior points. Near contours points appear and disappear so the change in these areas is not invertible. This is of course well-known in the field, and it is common to check for each point in the picture if the assumption that the mapping really is invertible. If it not is invertible the point is selected as a contour point candidate.

This approach gives quite indirect knowledge about what happens along contours. An alternative approach is to study the spatio-temporal patterns directly. By doing this we have direct information about what happens in the neighborhood of contour points. In the above discussed spatial case, our goal was to recognize a certain surface patch under a large class of viewing and lightening transformations and possibly also under other classes of transformations. Our corresponding goal for spatio-temporal vision is to be able to classify small regions of the spatio-temporal input patterns, preferably based on “natural” classes of events in the surroundings. Our first task is therefore to impose an appropriate spatio-temporal geometry on the visual input.

Time

As discussed earlier we consider the spatio-temporal visual input as a smooth scalar function over a small region in three-space, \( u \in C(U), U \subset \mathbb{R}^3 \). In the small spatio-temporal regions that biological seeing creatures operate in, there is no need for handling relativistic effects and we can use the concept of absolute time\(^4\). There is a temporal interval between any two points \( x, y \in U \) in spatial time, from this one can introduce a time function \( t : U \to \mathbb{R} \). The sets \( S(t_0) = \{ x | t(x) = t_0 \} \), are called planes of simultaneity. The space-time can be stratified in a sequence of planes of simultaneity, and be given coordinate systems that separates time and space, \( (t, x) \in U \subset \mathbb{R} \times \mathbb{R}^2 \). In time we consider an event in the same way, irrespective of when it happens, which calls for the group of 1-dimensional translations. In many cases, the time elapsed, is not important either, which calls for the scaling group. For longer time-sequences more general warping of the temporal domain could be useful, but for local properties we only care about translation and scaling, i.e. \( GL(1) \),

\[
t' = \lambda t + t_0,
\]

\(^4\)More elaborate discussions of the concepts given here can be found in e.g. Weyl (1952) and Friedman (1983).
CHAPTER 2. GEOMETRY OF VISUAL INFORMATION

should be a sufficient approximation.

Space-Time

From the consequences of absolute time, we conclude that we only want to allow for space-time transformations that never mix the planes of simultaneity.

Definition 2.4.1. General Galilean transformations on points in time-space, \((t, x) \in \mathbb{R} \times \mathbb{R}^n\), are transformations of the form,

\[(t', x') = (G(t), F_t(x)),\]

where \(F : \mathbb{R} \rightarrow \text{Diff}(\mathbb{R}^2), \ t \mapsto F_t\) and \(G : \text{Diff}(\mathbb{R})\).

We will focus our work on Affine Galilean transformations:

Examples 2.4.1. We list some affine Galilean transformations, in matrix form.

1. The Galilean affine groups, \(\Gamma A(n+1) \subset \Gamma A(n+1)\), acts on points in space-time \((t, x) \in \mathbb{R}^{n+1}\), in the following way:

\[
\begin{pmatrix}
 t' \\
 x'
\end{pmatrix} =
\begin{pmatrix}
 \lambda & 0 \\
 v & A
\end{pmatrix}
\begin{pmatrix}
 t \\
 x
\end{pmatrix} + a
\] (2.8)

where \(a, v \in \mathbb{R}^n, t \in \mathbb{R}, \lambda \in \mathbb{R}_+, \) and \(A \in \text{GL}(n)\) that only acts on space.

2. We can also consider less general action on space than the general linear group, e.g., if we put the restriction, \(A \in \mathbb{R} SO(n)\), in the above definition, we get the Galilean similarity group, \(\Gamma S(n)\).

3. The Galilean group, \(\Gamma_{n+1}\), has the form:

\[
\begin{pmatrix}
 t' \\
 x'
\end{pmatrix} =
\begin{pmatrix}
 1 & 0 \\
 v & R
\end{pmatrix}
\begin{pmatrix}
 t \\
 x
\end{pmatrix} + a
\] (2.9)

where \(a, v \in \mathbb{R}^n, t \in \mathbb{R}, \) and \(R \in \text{SO}(n)\). Each Galilean motion can be decomposed in a product of a spatial rotation:

\[
\begin{pmatrix}
 t' \\
 x'
\end{pmatrix} =
\begin{pmatrix}
 1 & 0 \\
 0 & R
\end{pmatrix}
\begin{pmatrix}
 t \\
 x
\end{pmatrix}
\] (2.10)

a spatio-temporal shear (constant velocity):

\[
\begin{pmatrix}
 t' \\
 x'
\end{pmatrix} =
\begin{pmatrix}
 1 & 0 \\
 v & I
\end{pmatrix}
\begin{pmatrix}
 t \\
 x
\end{pmatrix}
\] (2.11)

and a space-time translation:

\[
\begin{pmatrix}
 t' \\
 x'
\end{pmatrix} =
\begin{pmatrix}
 t \\
 x
\end{pmatrix} + a
\] (2.12)
2.4. SPATIO-TEMPORAL GEOMETRY

Table 2.1: The subgroups and their infinitesimal generators for the Galilean similarity group

<table>
<thead>
<tr>
<th>Transformation</th>
<th>( g_p(t, x, y) )</th>
<th>( v )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spatial</td>
<td></td>
<td></td>
</tr>
<tr>
<td>x-translation</td>
<td>((t, x + p_1, y))</td>
<td>(\partial_x)</td>
</tr>
<tr>
<td>y-translation</td>
<td>((t, x, y + p_2))</td>
<td>(\partial_y)</td>
</tr>
<tr>
<td>rotation</td>
<td>((t, \cos p_3 x + \sin p_3 y, \sin p_3 x - \cos p_3 y))</td>
<td>(y\partial_x - x\partial_y)</td>
</tr>
<tr>
<td>scaling</td>
<td>((t, p_4 x, p_4 y))</td>
<td>(x\partial_x + y\partial_y)</td>
</tr>
<tr>
<td>Temporal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>translation</td>
<td>((t + p_5, x, y))</td>
<td>(\partial_t)</td>
</tr>
<tr>
<td>scaling</td>
<td>((p_5 t, t, x, y))</td>
<td>(t\partial_t)</td>
</tr>
<tr>
<td>Spatio-temporal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>x-Galilean boost</td>
<td>((t, x + p_7 t, y))</td>
<td>(t\partial_x)</td>
</tr>
<tr>
<td>y-Galilean boost</td>
<td>((t, x, y + p_8 t))</td>
<td>(t\partial_y)</td>
</tr>
</tbody>
</table>

Defintion 2.4.2. The Galilean similarity group, \(\Gamma_S\), on \(\mathbb{R}^2 \times \mathbb{R}\), \((t, x)\) consists of a number of spatial motions: translation, rotation and scaling, a number of temporal motions: translation and scaling and a spatio-temporal motion: Galilean boost. The groups and their infinitesimal generators are gathered in table 2.4.

Spatio-Temporal Lie Algebra

The 1+1 dimensional Galilean similarity group, i.e. translation invariance in space and time, separate scaling in space and time and Galilean boost in space time, have the following set of infinitesimal generators,

\[
\gamma_S(2) = \mathfrak{l}(2) \cup \mathfrak{s}(1) \oplus \mathfrak{s}(1) \cup \gamma(1),
\]

where \(\gamma(1) = \{ \gamma = x_0 \partial_t \} \) is the Galilean boost that “skew” space-time and \(\mathfrak{s}(1) \oplus \mathfrak{s}(1)\) is a direct sum of the scaling generator in space and time respectively. The non-zero commutators are \([\partial_j, x_0 \partial_t] = \partial_t, [\partial_0, \gamma] = \partial_t, [x_0 \partial_t, \gamma] = \gamma\) and \([x_1 \partial_t, \gamma] = -\gamma\).

The 2 + 1 dimensional Galilean similarity group is generated by,

\[
\gamma_S(3) = \mathfrak{c}_S(2) \oplus \mathfrak{gl}(1) \cup \gamma(2)
\]

where \(\gamma(2) = \{x_0 \partial_1, x_0 \partial_2\}\).

The \(d + 1\) dimensional Galilean similarity group is generated by,

\[
\gamma_S(d + 1) = \mathfrak{c}_S(d) \oplus \mathfrak{gl}(1) \cup \gamma(d).
\]

Minkowskian Space

Hoffman (Hoffman 1966) and Caelli (Caelli, Hoffman & Lindman 1978) have proposed to use the 2+1 dimensional Lorentz group in the description of the human visual system. Experiments show that the perceived length of a moving object decreases as its velocity increases, and that there is a maximum velocity of perceived
CHAPTER 2. GEOMETRY OF VISUAL INFORMATION

movement. Artificial seeing systems cannot estimate velocity larger than the size of the receptive field per frame, due to the sampling theorem\(^5\) (Jahne 1993). This maximum velocity makes the Lorenz group appropriate also for artificial systems. It is however not obvious that effects such as length contraction are desirable in this area.

2.5 Invariance and Covariance

For small changes in camera position, the change between the pictures, around an interior point, can be described by an element in an appropriate transformation group. If we consider the pictures as functions over a certain $G$-space, this is an equivalence relation on the $G$-space.

**Definition 2.5.1.** Two functions, $f_1, f_2 : X \to Y$, over a $G$-space $(X, G)$ are said to be equivalent\(^6\) if

$$f_2(x) = f_1(g \cdot x), \quad x \in X$$

For our purposes, however this kind of equivalence is, for several reasons, too “strong” to be useful. First: in practice, the change between the pictures, is not completely reversible: some small amount of “information” appear and disappear. Second: even within the invertible part of the change, we can not hope to know all involved groups, but just the most important ones - the ones that are responsible for most of the change. Third: in spite of the seemingly simple form of the equivalence, it is computationally hard to check for equivalence, in a direct way. Despite these complications, humans are for example, in most cases able to recognize a surface patch from different directions. Obviously some kind of property must stay the same, irrespective of the change. This leads to the concept of invariance: instead of considering equivalence between pictures, one considers equivalence over some property of the picture. Such properties are called invariants. More generally:

**Definition 2.5.2.** A function $I : X \to Y$ is invariant w.r.t. the group $G$ that acts on $X$ iff for all $g \in G, x \in X$ $I(g \cdot x) = I(x)$.

The above mentioned problem with checking equivalence, can then be approached by finding and checking a complete set of invariants. An important task for a seeing system is therefore to be able to detect various invariant properties in the ambient optical array.

Maybe we should add a little to how adding invariants leads to better approximation of equivalence: If a group $G$ acts on a function space $F$, (in our case: pictures), each function in the space generates an orbit $G(f), f \in F$, and the function space will be partitioned in disjunct orbits, $F/G$. Each orbit is an equivalence class of functions. An invariant has obviously the same value for all functions within

\(^5\)We can do better with e.g. methods based on matching or Kalman filters, but this requires additional knowledge.

\(^6\)More specifically, the equivalence relation is a $G$-space auto-morphism.
an orbit, but in most cases, also for several different orbits. Thus, such an invariant generates a coarser partitioning of the function space. Using more invariants leads to a finer partitioning of the function space and thus a better approximation of equivalence.

In the current work we choose some $G$-spaces that, based on various arguments, are good first approximations of the visual input. We then try to find out how to measure some basic invariants in these spaces. A more intriguing question that biological vision systems have to solve, through learning or evolution, is how to find both symmetry groups that approximate those in the visual input, and invariants relative to these.

Felix Klein, in his famous Erlangen program, 1872, defined geometry as:

Given a manifoldness and a group of transformations of the same; to investigate the configurations belonging to the manifoldness with regard to such properties as are not altered by the transformation of the group.

(see e.g. Sharpe (1997)). From this perspective, we can see our endeavor, as the study of the (Klein) geometry\(^7\), or rather the differential geometry, (as we only consider local properties), of functions over certain $G$-spaces.

Sometimes we are not able to find invariants but still want families of function that are as compatible with the transformation group as possible. Such families of function are said to be covariant.

Definition 2.5.3. A family of functions $G \rightarrow (X \rightarrow Y)$, $g \mapsto f_g : X \rightarrow Y$ is covariant w.r.t. the group $G$ iff $g \cdot f_h = f_{g \cdot h}$, $g, h \in G$, i.e. $g \cdot f_h(x) = f_h(g \cdot x) = f_{g \cdot h}(x)$, $x \in X$ and contravariant w.r.t. $G$ iff $g \cdot f_h = f_{h \cdot g}$.

2.6 Infinitesimal Invariance

It is often useful to describe the infinitesimal effect of groups as this linearizes the problems and in most cases we are only interested in the local effects. The infinitesimal generator of $g_p$ at $x$ is:

$$v|_x = \partial_p g_p(x)|_{p = \text{id}},$$

and if $g_p$ acts of the function $f$,

$$\partial_p g_p \cdot f(x)|_{p = \text{id}} = \partial_p f(g_p(x))|_{p = \text{id}} = \partial_1 f(g_p(x))\partial_p g_p(x)|_{p = \text{id}} = v f(x).$$

\(^7\)Riemann geometry does not fit into this context, but there is a generalization, Cartan geometry, that includes both kind of geometries (Sharpe 1997).
CHAPTER 2. GEOMETRY OF VISUAL INFORMATION

Invariance and Covariance

One can also define infinitesimal invariance and covariance. We start with one-parameter Lie-transformation groups.

Proposition 2.6.1. A function $f$ is infinitesimally invariant w.r.t. the one parameter Lie transformation group $g_p$ iff $vf = 0$, where $v$ is the infinitesimal generator of the group.

Proof.

\[ 0 = \partial_p(g_p \cdot f(x) - f(x))|_{p=\text{id}} \\
= \partial_p(g_p(x))|_{p=\text{id}} \\
= vf(x). \]

Proposition 2.6.2. A family of functions $f_q$ are infinitesimally covariant w.r.t. a group $g_p$, $p,q \in G$ where $G$ is an one-dimensional abstract Lie group, iff $a(q)\partial_q f_q = vf_q$ where $a(q) = \partial_p(p \cdot q)|_{p=\text{id}}$.

Proof.

\[ 0 = \partial_p(g_p \cdot f_q(x) - f_p \cdot q(x))|_{p=\text{id}} \\
= v_q f_q(x) - \partial_1 f_p \cdot q(x)\partial_p(p \cdot q)|_{p=\text{id}} \\
= v_q f_q(x) - a(q)\partial_q f_q(x). \]

For $n$-parameter Lie groups the we get:

Proposition 2.6.3. A function $f$ is infinitesimally invariant w.r.t. the $m$-parameter Lie transformation group $g_p$ on a $n$-dimensional manifold iff $vf = 0, \forall i$, where $\{v_i\}$ span the vector space of infinitesimal generators and $\{p_i\}$ is the corresponding canonical parametrization of the group.

Proof.

\[ 0 = D_p(g_p \cdot f(x) - f(x))|_{p=\text{id}} \\
= D_p(f(g_p(x))|_{p=\text{id}} \\
= D_1 f(g_p(x)) \cdot D_p g_p(x)|_{p=\text{id}} \\
= D_x f(x) \cdot (D_p g_p(x)|_{p=\text{id}}) \\
= D_x f(x) \cdot b(x) \\
= (v_1 f(x), \ldots, v_m f(x)) \]

where $b(x) = D_p g_p(x)|_{p=\text{id}}, b_{ij}(x) = \partial_p g_{i,p_j}(x)|_{p_j=\text{id}}$ and $v_j = \sum b_{ij}(x)\partial_i$. 

2.6. INFINITESIMAL INVARIANCE

Proposition 2.6.4. An m-parameter family of functions \( f_q \) is infinitesimally covariant w.r.t. the m-parameter Lie transformation group \( g_p \) on a n-dimensional manifold iff \( \partial_q f_q = v_i f_q, \forall i \), where \( \{v_i\} \) spans the vector space of infinitesimal generators and \( \{p_i\} \) is the corresponding canonical parametrization of the group.

Proof.

\[
0 = D_p(g_p \cdot f_q(x) - f_{p \cdot q}(x))|_{p=\text{id}} \\
= D_x f_q(x) \cdot b(x) - (D_p f_{p \cdot q}(x)|_{p=\text{id}}) \\
= D_x f_q(x) \cdot b(x) - (D_1 f_{p \cdot q}(x) \cdot D_p(p \cdot q)|_{p=\text{id}}) \\
= D_x f_q(x) \cdot b(x) - D_q f_q(x) \cdot a(q) \\
= (v_1 f(x), \ldots, v_n f(x)) f(x) - D_q f_q(x)
\]

where \( a(q) = D_p(p \cdot q)|_{p=\text{id}} \) and as the parametrization is canonical \( a(q) = \text{id}. \)

We will need a generalization of covariance:

Proposition 2.6.5. An m-parameter family of functions \( f_q \in G \) is infinitesimally covariant w.r.t. the m’-parameter Lie transformation group \( g_p, p \in G’ \), \( m \leq m’ \) and the mapping \( h : G’ \times G \to G \) on an n-dimensional manifold iff \( D_x f_q(x) \cdot b(x) = D_q f_q(x) \cdot a(q) \), where \( a(q) = D_p(h(p, q))|_{p=\text{id}} \) and \( b(x) = D_p g_p(x)|_{p=\text{id}} \)

Proof.

\[
0 = D_p(g_p \cdot f_q(x) - f_{h(p, q)}(x))|_{p=\text{id}} \\
= D_x f_q(x) \cdot b(x) - (D_p h_{(p, q)}(x)|_{p=\text{id}}) \\
= D_x f_q(x) \cdot b(x) - (D_1 h_{(p, q)}(x) \cdot D_p(h(p, q))|_{p=\text{id}}) \\
= D_x f_q(x) \cdot b(x) - D_q f_q(x) \cdot a(q)
\]

Galilean Invariants

It can be shown that planes of simultaneity (constant time) are invariant and has Euclidean geometry, i.e. distances and angles are invariants.

The temporal distance between planes of simultaneity is invariant.

Definition 2.6.1 (Galilean Vector Space). A Galilean vector space \( \Gamma_{n+1} = T \oplus S \) is a direct sum of a one-dimensional temporal vector space over \( \mathbb{R} \), \( T \) and a n-dimensional spatial Euclidean vector space \( S \) together with two symmetric bilinear forms \((\cdot, \cdot)_T\) and \((\cdot, \cdot)_S\) such that:

\[
\begin{cases}
(x, x)_T > 0 & \text{if } x \in T, \\
(x, y)_T = 0 & \text{if } x \in S \lor y \in S.
\end{cases}
\]

\((x, y)_S\) is the ordinary (Euclidean) scalar product and it is only defined if \( x, y \in S \).

There are two norms, defined as \( \|x\|_T = \sqrt{(x, x)_T} \) and \( \|x\|_S = \sqrt{(x, x)_S} \).
Galilean frames A Galilean coordinate system is an affine coordinate system where the spatial part consists of ON-coordinate system.

Definition 2.6.2. A Γ-base is a set \( \{ b_i \}_{i=0,...,n} \subset \Gamma_{n+1} \) such that \( 0 \neq b_0 \in T \) and \( \{ b_i \}_{i=1,...,n} \) is a base in \( S \). A Γ-base \( \{ e_i \} \) is called orthogonal if \( \langle e_i, e_j \rangle_S = 0 \) for \( i \neq j \) and orthonormal if it is orthogonal and \( \| e_0 \|_T = 1 \) and \( \| e_i \|_S = 1 \), \( i = 1,\ldots,n \).

Classification Each \( \Gamma_3 \) Galilean motion is a skew screw motion, i.e. a rotation in the spatial planes around a line that cuts the spatial planes, followed by a translation parallel with the line.

2.7 Generalizations

We end this chapter by showing how the visual geometry spaces can be defined on more general spaces than Euclidean spaces. These results are not used in the rest of the thesis.

Homogeneous Manifolds

Definitions 2.7.1. Let \( \sigma \) be an action of the Lie group \( G \) the manifold \( M \), then for each \( x \in M \), \( \sigma(G)x \) is called an orbit.

\[ G_x = \{ g \in G \mid \sigma(g)x = x \} \]

is called the stabilizer (or isotropy subgroup) at \( x \). A Lie group action \( \sigma \) is transitive if for any \( x, y \in M \) there is a \( g \in G \) such that \( y = \sigma(g)x \). In this case \( M \) is said to be a homogeneous manifold of the group \( G \).

Being in the same orbit is an equivalence relation on \( M \). The group action is obviously transitive within each orbit. A transitive action has only one orbit: the whole manifold. It can be shown for any base point \( x \in M \), that the map:

\[ f_x : G/G_x \to M, \quad gG_x \mapsto gx, \]

is a \( G \)-space isomorphism. It can also be shown that, if two points, \( x, y \in M \) are on the same orbit, i.e. there exists an element \( g \in G \) such that \( y = gx \), then the points have conjugate stabilizers, \( G_y = a_y(G_x) \). Thus a homogeneous manifold \( M \) of the group \( G \) is isomorphic with \( G/G_x \) for any of its stabilizer. A homogeneous \( G \)-space can be completely described as a pair \( (G, H) \), where \( H \) is a closed subgroup of \( G \) and \( G \) acts on \( G/H \) by means of left translation.

Examples 2.7.1. The orbit of \( \text{SO}(n) \) acting on a point \( x \in \mathbb{R}^n \) is the \( n-1 \)-sphere with radius \( |x| \). The orbit of \( E(n) \) and its supergroups (e.g. \( \text{ES}(n) \) and \( \text{GA}(n) \)) acting on any \( x \in \mathbb{R}^n \), is all of \( \mathbb{R}^n \). Thus the action of \( E(n) \), \( \text{ES}(n) \) and \( \text{GA}(n) \) are transitive, while the action of \( \text{SO}(n) \) is non-transitive. The stabilizers at \( 0 \in \mathbb{R}^n \) for the above mentioned subgroups of the general affine group, can easily be seen to be:
GA\(n\)\(_0\) = GL\(n\), ES\(n\)\(_0\) = \(\mathbb{R}_+\) SO\(n\) and E\(n\)\(_0\) = SO\(n\), respectively. Their G-spaces can therefore be described by the pairs: (GA\(n\), GL\(n\)), (ES\(n\), \(\mathbb{R}_+\) SO\(n\)) and (E\(n\), SO\(n\)).

**G-Bundles**

Now, after reviewing Lie group action on manifolds, we want to extend these actions to function spaces, or rather, smooth bundles, (Onishchik 1993).

**Definition 2.7.2.** Let \(G\) be a Lie group, then \(\pi : E \to B\) is a \(G\)-bundle if \(\pi : E \to B\) is a smooth bundle and \(G\), acts on both \(E\) and \(B\) in such a way that \(\pi\) is an equivariance. I.e. for \(u \in E, g \in G\), \(\pi(gu) = g\pi(u)\).

**Definition 2.7.3.** A \(G\)-bundle morphism between the \(G\)-bundle \(E \xrightarrow{\pi} B\), and the \(G'\)-bundle \(E' \xrightarrow{\pi'} B'\), is a pair \((u, f)\) of two continuous maps \(E \xrightarrow{u} E'\), and \(B \xrightarrow{f} B'\), where \(f\) is a \(G\)-space morphism, such that the following diagram commutes.

\[
\begin{array}{ccc}
E & \xrightarrow{u} & E' \\
\downarrow{\pi} & & \downarrow{\pi'} \\
B & \xrightarrow{f} & B'
\end{array}
\]

**Example 2.7.2.** A smooth product bundle \((B \times F, \text{proj}_1, B, F)\), can be given \(G\)-bundle structure, with the following action:

\[
g(x, y) = (gx, gy), \quad g \in G, x \in B, y \in F,
\]

since

\[
\text{proj}_1(g(x, y)) = \text{proj}_1(gx, gy) = gx = g\text{proj}_1(x, y).
\]

This kind of action can be considered as a fiber preserving transform if we identify a section of the bundle with the graph of some map \(B \to F\). In the same sense, actions of the form: \(g(x, y) = (gx, y)\), corresponds to base transforms.

**Example 2.7.3.** Let GL\(n\) act on \(\mathbb{R}^n\), this action can be extended to an action on functions \(f : \mathbb{R}^n \to \mathbb{R}\), by

\[
gf(x) = f(gx), \quad g \in \text{GL}(n), x \in \mathbb{R}^n.
\]

On the section \(s \in C^\infty(\mathbb{R}^n \times \mathbb{R}, \xrightarrow{\text{proj}_1} \mathbb{R}^n)\), this corresponds to the action base action \(g(x, y) = (gx, y), \quad (x, y) \in s\). We can also construct a volume preserving action of GL\(n\) on \(f : \mathbb{R}^n \to \mathbb{R}\) by,

\[
gf(x) = \det(g)f(gx).
\]

Or on the corresponding bundle section:

\[
g(x, y) = (gx, \det(g)y).
\]
Definition 2.7.4. A \textit{homogeneous} $G$-bundle, is a $G$-bundle where the base space is a homogeneous manifold $G/H$, where $H \subset G$.

Visual Input Model

Let us now return to the vision case and summarize our model of spatial visual input. The visual input is a linear function space over a two-dimensional manifold. In terms of smooth bundles, we can say that the visual input is a section of a bundle $M^2 \times \mathbb{R} \xrightarrow{\text{proj}_1} M^2$, where the fiber is a linear space. Such bundles are called a vector bundles (see Appendix B.4). We also have Lie groups that acts on the input, which can be described in terms of vector $G$-bundles, ($G$-bundles that also are vector bundles). We are mainly interested in transitive group action, which leads to homogeneous vector $G$-bundles. To summarize: our (preliminary) view is that the visual input is a homogeneous vector $G$-bundle.
Chapter 3

Geometrical Measurement

3.1 Introduction

In this chapter we will investigate how to measure and represent the spatio-temporal visual input in a way that is appropriate for visually guided agents like man, higher mammals and possibly robots. Such an agent uses optic information for a wide range of different tasks, such as knowing about its motion relative to environment, finding and recognizing objects, finding out the three-dimensional structure of environment and driving attention just to mention a few.

In the previous chapter we proposed that the visual input signal - a picture or a movie, can be described as a linear space of scalar functions over a two- or three-dimensional manifold together with an appropriate set of Lie transformation groups. A basic visual ability, the one we pursue in this work, is to separate the visual information in variant and invariant aspects with respect to important Lie transformation groups.

As a first step in the various visual competences the observer must make measurements from the ambient light and embody the measurements, i.e. make them available for later processing steps. The part of the visual system that make the first processing step is called the visual front end. If the visual front end should serve as input to a wide range of visual competences it must represent information in a rather general form, in such a way that it is compatible with the structure of the input signal. We will follow the general principle that: the visual front-end should embody as much of the structure from the visual input as possible. Koenderink, Kappers & van Doorn (1992) suggest that the visual front end should be a syntactical transformer, with no semantic interpretation of the data in terms of properties of the environment. It works in a bottom up fashion, i.e. it depends on the visual input and not on semantic hypotheses from higher-level computational stages.

The visual front-end should also focus on local properties, as non-local properties have quite complicated structure (Koenderink et al. 1992). To be able to measure
non-local properties, a front-end system would have to have a method to compare distant points i.e. a connection which is far more complicated than just having local competence. It is also much more complicated to learn non-local properties, as the correlation structure between distant points is much more complicated to establish than the correlation structure in a neighborhood. From this it seems reasonable to assume that non-local properties are more appropriate to handle in “high”-level parts of a visual system, and that is beyond our current scope.

At first sight it might seem natural to measure the value of the movie in each point. However it is impossible to know the value in a point: a measurement must collect light from a non-vanishing region of space-time, an extended point. To make the measurement as point like as possible it should still be localized and positivity preserving. As the visual input has linear structure, we require the measurements to have linear structure as well. The consequences of these requirements (except for positivity preservation) will be discussed in Section 3.2.

The measurement should also embody the structure of the input signal. This means that it should have two or three dimensional manifoldness, although possibly in another coordinate system. It also means that it should respect the Lie group structure of the input. The latter requirement is trivially fulfilled for point measurements, but as soon as the measurement has volume it also has shape and this shape might not be invariant to the group of transformations. For many Lie groups, there are no local invariant measurement operators. In these cases we will use a family of measurement operators that is covariant with respect to the group instead. (Section 3.3). Ideally the family of measurements from the family of measurement operators should have the same Lie group structure as the signal. In many cases this is not possible and we have to be satisfied with semi-group structure as we will see in section 3.4.

To simplify calculations and to enable measurements on e.g. bounded spaces we derive infinitesimal conditions for geometrical measurement in Section 3.5. In Section 3.6 we will study the consequences of the positivity preservation requirement. Finally in Section 3.7 we give a infinitesimal characterization of a $G$ scale-space.

This and the following chapter is an extension of Fagerström (2007).

### 3.2 Extended Point

In the previous discussion we have thought about spatio-temporal signals as smooth functions from space-time to scalar values, $u \in \Sigma \equiv C^\infty(\Omega), \Omega \subset M$. One might then be tempted to view $u(x)$, as a measurement at $x \in \Omega$. However, measurement of the instantaneous value of a signal at a certain point is physically impossible, each measurement must take a non vanishing amount of time and cover a non vanishing area of space$^1$. Furthermore, assuming smoothness of the signal might

---

$^1$This section is to a large extent a review of the measurement framework developed by Florack (1997).
be unrealistic. On the other hand, the measurement process, must impose some kind of coherence on the world, otherwise, an embodiment of the visual input using finite resources is impossible. To make the concept of local measurement plausible, the measurement must allow for at least a few derivatives. It can be shown that a $C^r$-function, $1 \leq r \leq \infty$, can be arbitrary well approximated, by a $C^\infty$-function, (Hirsch 1976), and we can therefore as well require measurement to be smooth.

From the above reasoning we make a new trial to define signals and measurement: we let signals be arbitrary functions, $\Sigma \equiv \Omega \to \mathbb{R}$, and introduce a measurement operator, $\Phi : \Sigma \to C^\infty(M)$.

**Linearity**

As measurement is supposed to respect the structure of the input and as the input is linear, measurement should be linear as well,

$$\Phi(c_1u_1 + c_2u_2) = c_1\Phi(u_1) + c_2\Phi(u_2),$$

where $c_1, c_2 \in \mathbb{R}, u_1, u_2 \in \Sigma$. We can also define a point measurement as, for all $x \in M$, $\Phi_x : L(\Sigma \to \mathbb{R}), u \mapsto \Phi(u)(x)$. Point measurements are thus functionals on signal space, and we can define the space of point measurements, $\Delta$, as the topologically dual space to the signal space, $\Delta \equiv \Sigma'$ (Treves 1967). A conceptual drawback with this definition is that the space of measurement devices is defined in terms of the signal space, and the only way to know something about the signal space is to actually make measurements. We seem to have started from the wrong space. There are also some technical problems with the definition, (see Florack (1997) for details). The space of point measurements becomes to “large”, and includes the kind point sized measurement that we wanted to avoid.

**The Signal as a Distribution**

Another and more realistic approximation of the situation is to shift focus to the part of the measurement situation that we can control, the measurement device. We represent the measurement devices by smooth functions

$$\phi : \Delta \equiv M \to \mathbb{R}$$

where $M$ is an $n$-dimensional manifold and the signal as a function from measurement devices to scalars

$$u : \Sigma \equiv \Delta \to \mathbb{R} = \Delta'.$$

The signal is considered as a “black box” that is probed by a measurement device which results in a measurement $u(\phi)$. We also require a signal $u \in \Sigma$ to be linear with respect to the measurement devices:

$$u(c_1\phi_1 + c_2\phi_2) = c_1u(\phi_1) + c_2u(\phi_2) \quad c_1, c_2 \in \mathbb{R}, \phi_1, \phi_2 \in \Delta$$
i.e. \( u \) is a functional. Furthermore, we require that the signal is continuous with respect to an appropriate topology on the space of measurement devices. At first sight linearity may seem to be a quite restrictive requirement on the signal \( u \), however one may as well consider it as requirements on the choice of measurement devices relative to the signal. The numerical labeling of the output levels from the measurement, i.e. the coordinate system of the output from a measurement, is a design choice, and we are free to choose any diffeomorphism that makes the system linear. The consequence of the linearity requirement is that measurements shall be possible to combine and the consequence of the homogeneity requirement (\( u(a\phi) = au(\phi) \)), is that all intensity levels are treated in the same way.

The mathematical structure described above is developed in great detail under the name of distribution theory, where, what we called signals above, are denoted distributions and measurement devices are denoted test functions. The modern theory of distributions was introduced during the 1940’s by Laurent Schwartz. Florack has advocated (Florack 1997) its use for representing images.

**Point Measurement**

An image is defined as a distribution \( u \in \Sigma(\Omega) \) over some connected subset of an Euclidean base manifold \( \Omega \subset M, M = \mathbb{R}^d \ni x = (x_1, \ldots, x_d) \) for spatial images and \( M = \mathbb{R}^{d+1} \ni x = (x_0, x_1, \ldots, x_d) \) for spatio-temporal images, where \( x_0 \) is a coordinate in the temporal direction.

Following standard motivations from scale-space theory (see (Koenderink et al. 1992, Florack 1997)) we define point measurement,

\[
u = \Phi f,\]

as:

**Definition 3.2.1.** Let \( f, g \in \Sigma \) and \( \alpha, \beta \in \mathbb{R} \), then \( \Phi : \Sigma \rightarrow C^\infty(\Omega) \) is a point measurement operator if it fulfills:

- **linearity** \( \Phi(\alpha f + \beta g) = \alpha \Phi(f) + \beta \Phi(g) \)
- **gray level invariance** \( \|\Phi f\|_{L_1} = \|f\|_{L_1} \)
- **positivity** \( f \geq 0 \Rightarrow \phi f \geq 0 \)

**point** There is a sequence of operators \( \mathbb{R}_+ \ni s \mapsto \Phi_s \) s.t. \( \lim_{s \to 0} \Phi_s f = f \)

From the linearity requirement it can be shown that the point measurement operator on a function \( f \in L_1(\Omega) \) is an integral operator,

\[
\Phi f(x) = \int_{\Omega} \phi_x(y) f(y) dy = \langle f, \phi_x \rangle,
\]
3.3 Covariance

with a smooth point measurement kernel $\Omega \ni x \mapsto \phi_x \in C^\infty(\Omega, \mathbb{R}) = \Delta$. More generally for $f \in \Sigma$:

$$\Phi f(x) = f(\phi_x).$$

Differentiation of the possibly non-smooth images can be defined in terms of the differentiable kernel functions,

$$\langle \partial^\alpha f, \phi \rangle = \langle f, (-1)^{|\alpha|} \partial^\alpha \phi \rangle,$$

where $\alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbb{Z}_+^d$ is a multi index, $\partial^\alpha = \partial_1^{\alpha_1} \cdots \partial_d^{\alpha_d}$, $|\alpha| = \alpha_1 + \cdots + \alpha_d$ and $\alpha! = \alpha_1! \cdots \alpha_d!$.

From gray level invariance we get that the measurement kernel must possess unit volume:

$$\int \phi_x = 1. \quad (3.2)$$

The positivity requirement implies a positive kernel,

$$u \geq 0 \Rightarrow u(\phi_x) \geq 0 \quad (3.3)$$

and the point property means that the measurement kernel approaches a Dirac pulse.

$$\lim_{x \to 0} \phi_x = \delta, \quad (3.4)$$

3.3 Covariance

Image measurement should respect basic symmetries in the world that is measured. These symmetries can be described in terms of Lie groups acting on images. Let $g \in G$, where $G$ is a Lie group, then the group act on the base space as

$$g \mapsto g \cdot x = T_g x,$$

where $T_g : \text{Diff}(\Omega)$, and on functions like

$$T_g f(x) = f(g \cdot x).$$

A group action on an image can be transfered to a group action on the kernel by,

$$\langle T_g f, \phi \rangle = \langle f, g \cdot \phi \rangle = \langle f, J^{-1}_{g^{-1}} \circ \phi \circ g^{-1} \rangle.$$  

Ideally we would like point measurement operators that are invariant,

$$\forall g \in G : \Phi T_g = \Phi,$$

w.r.t. the group but for many groups that is not possible. I.e. if we look at how the scaling group acts on the kernel of the operator only a constant function or a derivative of the Dirac “function” are invariant kernels, but non of those fulfill our
axioms. We have to be content with requiring covariance, i.e. we have a family $H \ni h \mapsto \Phi_h$ of measurement operators (at this stage $H$ is just a set, we will attach more structure to it later), fulfilling,

$$T_g \Phi_h = \Phi_{g \cdot h} T_g. \quad (3.5)$$

**Definition 3.3.1.** We call such a family of point measurement operators a $G$-covariant point measurement space.

From the definition we prove:

**Proposition 3.3.1.** The function $G \times H \ni (g, h) \mapsto g \cdot h = \sigma(g, h) \in H$ is a Lie group homomorphism in the first argument, $\forall g_1, g_2 \in G$, and $G \ni g \mapsto \sigma_g : H \to H$ a Lie group action on the set $H$,

$$\sigma_{g_1, g_2} = \sigma_{g_1} \circ \sigma_{g_2}.$$  

**Proof.** From the covariance requirement: $T_g \Phi_h = \Phi_{\sigma_g(h)} T_g$, we have:

$$T_{g_1 g_2} \Phi_h = \Phi_{\sigma_{g_1 g_2}(h)} T_{g_1 g_2},$$

We also have:

$$T_{g_1 g_2} \Phi_h = T_{g_1} T_{g_2} \Phi_h = T_{g_1} \Phi_{\sigma_{g_2}(h)} T_{g_2} = \Phi_{\sigma_{g_1} \circ \sigma_{g_2}(h)} T_{g_1 g_2},$$

which proves the proposition. \qed

As a smaller set of measurement operators is easier to manage we prefer measurement spaces that are invariant to as large subgroup as possible.

We will require the Lie group $G$ to act transitively on the base space $\Omega$, i.e. for all $x, y \in \Omega$ there is a $g \in G$ s.t. $y = g \cdot x$. Furthermore we will restrict our attention to the case where the transitive action is the translation group. We will also require that $G$ has a scaling group as subgroup.

### 3.4 Cascade Property

Now we have a family of measurement devices $H \ni h \mapsto \Phi_h$ that is covariant under the chosen Lie group. But we still don’t know how the actual measurements, $\Phi_h u$ are related. If we consider the result of a measurement a signal that also can be measured it is natural to require the family of measurements to be closed under composition,

$$\Phi_{h_1} \Phi_{h_2} u = \Phi_{h_1 \cdot h_2} u, \quad (3.6)$$
3.4. CASCADE PROPERTY

where

\[ H \times H \ni (h_1, h_2) \mapsto h_1 \cdot h_2 \in H, \]

is an abstract Lie semigroup. Although there is a rich modern theory about Lie semigroups, it will for our needs be enough to consider Lie semigroups \( H \subset G \) that are subsets of a Lie group \((G, \cdot)\) and closed under composition but not necessarily under inversion. A simple example is \((\mathbb{R}_+, +)\).

From (3.6) we can see that \( \{ \Phi_h | h \in H \} \) is an operator semigroup and furthermore as \( H \ni h \mapsto \Phi_h \) is a Lie semigroup homomorphism, the operator semigroup must be a Lie semigroup as well.

Another motivation for requiring the space of measurement operators to be a semi group is that we want measurements of nearby parameters to be comparable. This means that for each \( h \in H, f \in \Sigma, \) a one parameter subgroup \( G_1 \subset G \) will generate a one-parameter family of measurements

\[ G_1 \ni g \mapsto \Phi_{g \cdot h} f. \]

Differentiability in the one parameter family of measurements along the parameter group means that the one parameter families

\[ G_1 \ni g \mapsto \sigma(g, h) \in H_1 \subset H \]

and

\[ H_1 \ni h \mapsto \Phi_h \]

must be Lie semi-groups in turn.

Covariant Semi-Groups

Combining \( G \)-covariance with the cascade property we finally state:

**Definition 3.4.1.** A \( G \) scale-space is a minimal semigroup of \( G \)-covariant point measurements, i.e. a family of operators \( H \ni h \mapsto \Phi_h : L(\Sigma, C^\infty(\Omega)) \) fulfilling

\[ T_g \Phi_h T_{g^{-1}} = \Phi_{g \cdot h}, \quad (3.7) \]

a slight reorganization of (3.5) to emphasize how \( G \) act on the semi group of operators and (3.6), and where \( G \) is an abstract Lie group and \( H \) an abstract Lie semigroup.

We have seen (Proposition 3.3.1) that

\[ G \times H \ni (g, h) \mapsto g \cdot h = \sigma(g, h) \in H \]

is a group homomorphism in the first argument. With \( H \) an Lie semigroup, it is also a semigroup homomorphism in the second argument.
Proposition 3.4.1. The function $G \times H \ni (g,h) \mapsto \sigma_g(h)$ is a Lie semi-group homomorphism in the second argument, $\forall h_1, h_2 \in H, \forall g \in G,$
\[ \sigma_g(h_1 h_2) = \sigma_g(h_1) \cdot \sigma_g(h_2). \]

Proof. From the covariance requirement we get:
\[ T_g \Phi_{h_1 h_2} = \Phi_{\sigma_g(h_1 h_2)} T_g, \]
and also:
\[ T_g \Phi_{h_1 h_2} = T_g \Phi_{h_1} \Phi_{h_2} = \Phi_{\sigma_g(h_1)} T_g \Phi_{h_2} = \Phi_{\sigma_g(h_1)} \Phi_{\sigma_g(h_2)} T_g = \Phi_{\sigma_g(h_1) \cdot \sigma_g(h_2)} T_g, \]
which proves the proposition.

When the homomorphism is trivial in the first argument we have an invariant operator,
\[ T_g \Phi_h T_{g^{-1}} = \Phi_h. \]

3.5 Infinitesimal Generators

We continue by deriving infinitesimal conditions for a $G$ scale-space. There are several advantages in doing that. It is more realistic in the sense that we mainly care for local symmetries. While we certainly want to measure an object subtending 2 visual degrees moving 2 visual degrees/second and an object subtending 4 visual degrees moving 1 visual degree/second in similar way, we don’t require the system to measure photons that pass by. By using infinitesimal constraints we can also use the theory with more general boundary conditions, e.g. for a bounded sensor. Furthermore it will simplify calculations as it linearizes the problem.

Lie Wedge

For a Lie semigroup $H$ the infinitesimal object is called a Lie wedge,
\[ \mathcal{L}H = \mathfrak{h}, \]
and is a closed cone in a Lie algebra. The properties for Lie algebras mentioned above also holds for Lie wedges with the restriction that it is not necessary for both $[v, w]$ and $[w, v]$ for $v, w \in \mathfrak{h}$ to be members of the wedge (Hille & Phillips 1957). We will only consider Lie semigroups that are generated by their Lie wedge, i.e.
\[ \exp(\mathbb{R}_+ \mathfrak{h}) = H. \]
3.5. INFINITESIMAL GENERATORS

Covariant Generators

From a $G$ scale-space, using
\[ \Phi(h)(x) = \Phi_h(x), \]
for emphasizing the dependency on the semigroup $H$, we get a corresponding infinitesimal operator
\[ B = d\Phi(e) : \mathfrak{h} \to L(\Sigma, \Sigma), \]
\[ \mathfrak{h} \ni w \mapsto B_w = B(w). \]
Choosing a base \( \{w_1, \ldots, w_m\} \subset \mathfrak{h} \) for the Lie wedge a corresponding base, \( \{B_1, \ldots, B_m\} \) where $B_k = B_{w_k}$, of the Lie wedge of infinitesimal generators for the semi group is given.

Taking the differential w.r.t. the semi-group $H$ for the covariance equation (3.7), we get,
\[ T_g B_u T_g^{-1} = B_{c_g(u)}, \] (3.8)
where we get $G \ni g \mapsto d\sigma_g = c_g : \mathfrak{h} \to \mathfrak{h}$ from $G \ni g \mapsto \sigma_g : H \to H$. Such an operator is called a covariant tensor operator (Barut & Raczka 1986), in mathematical physics. In coordinate form we have
\[ T_g B_u T_g^{-1} = \sum_l c_{g,kl} B_l, \] (3.9)
where
\[ c_g(w) = c_g(\sum_k b_k w_k) = \sum_k b_k c_g(w_k) \]
and further by letting
\[ c_g(w_k) = \sum_l c_{g,kl} w_l, \]
we have
\[ c_g(w) = \sum_{kl} b_k c_{g,kl} w_l. \]

From the Lie wedge homomorphism,
\[ w = \sum_j b_j w_j, \]
implies
\[ B_w = \sum_k b_k B_k, \]
and
\[ w = [w_k, w_l] \]
implies
\[ B_w = [B_k, B_l]. \]
A Lie algebra (or wedge) homomorphism onto a Lie algebra (or wedge) of linear operators is called a Lie algebra (or wedge) representation. So the scale space generators $\mathfrak{h} \ni w \mapsto B_w$ is a Lie wedge representation.

Taking the differential of eq 3.8 w.r.t. the Lie group we get:

$$[A_v, B_w] = B_{C(v)(w)}, \quad \text{(3.10)}$$

where $A_v$ is an infinitesimal generator of the group, $A = dT(e) : \mathfrak{g} \rightarrow L(\Sigma, \Sigma)$, $\mathfrak{g} \ni v \mapsto A_v = A(v)$. And where the differential of $G \ni g \mapsto C_g \in \mathfrak{h}$ becomes $C = d(g \mapsto C_g)(e)$, $\mathfrak{g} \ni v \mapsto C_v = C(v) : \mathfrak{h} \ni h$, and $\mathfrak{h} \ni w \mapsto (g \ni v \mapsto C(v)(w))$ is a Lie algebra representation in the first argument and a Lie wedge representation in the second. In coordinates, with a base $\{v_1, \ldots, v_n\} \subset \mathfrak{g}$ for the $\mathfrak{g}$ and a corresponding base $\{A_1, \ldots, A_n\}$ where $A_j = A_{v_j}$ for the Lie algebra action on the base space, we get:

$$[A_j, B_k] = \sum_l C_{j,kl} B_l, \quad \text{(3.11)}$$

where

$$C(v_j)(w_k) = C_j(w_k) = \sum_l C_{j,kl} w_l.$$

We denote $C$, the covariance tensor. For an infinitesimally invariant operator we have,

$$[A_v, B_w] = 0.$$

Note that requiring an operator $B_w$ to be infinitesimally invariant with respect to an action $A_v$ means for the covariance tensor that $C(v)(w) = 0$.

Now we take a look at the semigroup generated of a set of covariant tensor operators. A one-parameter Lie semigroup is generated by an element in the corresponding Lie wedge $w \in \mathfrak{h}$ by

$$\mathbb{R}_+ \ni s \mapsto \exp(sw) = h_w(s) \in H_w \subset H.$$

Let $u : H \times \Omega \rightarrow \mathbb{R}$ and set

$$\mathfrak{h} \ni w \mapsto u_w(s, x) = u(h_w(s), x),$$

$$\begin{cases} 
\partial_s u_w = B_w u \\
\lim_{s \to 0} u_{w}(s, x) = f(x), \quad \text{where } f \in \Sigma.
\end{cases} \quad \text{(3.12)}$$

For an abstract Cauchy problem, like (3.12), the solution can be described in terms of a semigroup

$$u_w(s, x) = \exp(sw) f(x) = \Phi_{h_w(s)} f(x).$$

And given that we have chosen semigroups $h \in H$, such that $h = \exp(sw)$ for some $s \in \mathbb{R}_+$, $w \in \mathfrak{h}$, we have (Hille & Phillips 1957),

$$u(h, x) = \Phi_h f(x). \quad \text{(3.13)}$$
Reconnecting to the $G$ scale-space axioms, Definition 3.4.1, the Cauchy problem (3.12) for a $g$ tensor operator obviously generates a $G$-covariant semigroup. Furthermore, looking at Definition 3.2.1, the semigroup will be linear as long as the infinitesimal generator $B_w$ is independent of the evolution parameter ($s$ in (3.12)). Also the point property follows from the boundary condition in (3.12). What is left to give an infinitesimal characterization of is the positivity and gray level invariance properties. But before we do that we will find a more explicit description of the semigroup generators.

**Pseudo-Differential Operators**

For being able to continue we must be more concrete about the form of the operators in (3.10). The infinitesimal generators of the transformation groups are on the form:

$$
\sum_{j \leq n} a_j(x) \partial_j,
$$

where $a_j : M \to \mathbb{R}$. In earlier work in scale-space theory the infinitesimal generator have been the inverse Fourier transform of some smooth function. By using pseudodifferential operators, $\Psi DO$, (Egorov & Shubin 1994) we get a large enough class of operators to embed both of them (actually this class includes all continuous linear maps $S(\mathbb{R}^n) \to S'(\mathbb{R}^n)$). To motivate the definitions of $\Psi DO$’s we start by taking the Fourier transform of (3.14). We will use $D_j = i^{-1} \partial_j$ to simplify the expressions,

$$
A = \sum_{|\alpha| \leq m} a_\alpha(x) D^\alpha,
$$

and applying it on,

$$
u(x) = (2\pi)^{-n} \int e^{ix \cdot \xi} \tilde{u}(\xi) d\xi,
$$

where

$$
\tilde{u}(\xi) = \int e^{-ix \cdot \xi} u(x) dx.
$$

we have

$$
Au(x) = (2\pi)^{-n} \int e^{ix \cdot \xi} a(x, \xi) \tilde{u}(\xi) d\xi,
$$

where

$$
a(x, \xi) = \sum_{|\alpha| \leq m} a_\alpha(x) \xi^\alpha,
$$

and is called the symbol of $A$ and the corresponding operator is denoted $a(x, D)$. In the theory of pseudodifferential operators more general symbols $a : C^\infty(\Omega, \mathbb{R}^n)$ are used. They are divided into different classes depending on how fast the arguments grows (Egorov & Shubin 1994), but in the current context the symbols will be constrained by covariance and positivity to nice enough symbol classes, with well
known properties. So we will treat the pseudodifferential operators as a formal algebra during the calculations.

The composition of two symbols is:
\[
c(x, \xi) = a(x, \xi) \circ b(x, \xi)
\]
(3.18)
\[
= \sum_{\alpha} \frac{1}{\alpha!} [\partial_\xi^\alpha a(x, \xi)] \cdot [D_\xi^\alpha b(x, \xi)],
\]
(3.19)
and the commutator determines a Lie algebra structure on the symbols.

In this paper we will only consider translation invariant scale-spaces which further restricts the form of the symbol. A base for the Lie algebra of translations is,
\[
t(n) = \{\partial_1, \ldots, \partial_n\},
\]
(3.20)
and the corresponding symbols are, \(\partial_j \sim i^{-1} \xi_j\).

**Lemma 3.5.1.** Translation invariant symbols are position independent
\[
b(x, \xi) = b(\xi).
\]
(3.21)

**Proof.** An infinitesimally translational invariant symbol \(b\) fulfills,
\[
[i^{-1} \xi_j, b(x, \xi)] = 0, \quad j \leq n.
\]
Leading to,
\[
\partial_{\xi_j} b(x, \xi) = 0, \quad j \leq n,
\]
i.e. the symbol is independent of \(x\).

We can also see that any two translationally invariant symbols \(b_i(\xi), b_j(\xi)\) commutes, \([b_i(\xi), b_j(\xi)] = 0\), which means that cones of translation invariant symbols automatically becomes Lie wedges.

### 3.6 Positivity and Gray Level Invariance

As mentioned above, a positive translation invariant linear operator has a positive kernel. It can be shown that a kernel \(\Omega \ni x \mapsto \phi(x)\), is positive iff its Fourier transform is *positive definite*, i.e. for each \(k \in \mathbb{N}\) and each set of \(\xi_1, \ldots, \xi_k \in \mathbb{R}^d\) the matrix \((\hat{\phi}(\xi_i - \xi_j))_{i,j=1,\ldots,k}\) is positive Hermitian, (e.g. (Jacob 1996)). Furthermore for kernels that generate a semigroup \(\mathbb{R}^d \ni s \mapsto e^{-sb(\xi)}\), their Fourier transform is \(\hat{\phi}(\xi) = e^{-sb(\xi)}\), where \(\mathbb{R}^d \ni \xi \mapsto b(\xi) \in \mathbb{C}\).

**Definition 3.6.1.** A symbol \(\mathbb{R}^d \ni \xi \mapsto b(\xi) \in \mathbb{C}\) is *negative definite* if the matrix \((b(\xi_i) + b(\xi_j) - b(\xi_i - \xi_j))_{i,j=1,\ldots,k}\) is positive Hermitian for any choice of \(k \in \mathbb{N}\) and each set of \(\xi_1, \ldots, \xi_k \in \mathbb{R}^d\).
3.7. INFINITESIMAL G SCALE-SPACE

A kernel is positive definite iff its symbol is negative definite (Jacob 1996). For negative definite symbols \( b(\xi), b(0) \geq 0 \) and they have at most quadratic growth at \( \infty \) i.e. \( |b(\xi)| \leq k_b(1 + |\xi|^2) \) for some \( k_b \in \mathbb{R}_+ \).

For gray level invariance we use an adaption of a theorem from (Jacob 1996).

**Theorem 3.6.1.** For a gray level invariant semigroup \( \Phi_s \) generated by the \( B \) with symbol \( b \) the following holds:

1. \( \Phi_s 1 = 1 \) for all \( s \geq 0 \)
2. \( B 1 = 0 \)
3. \( b(0) = 0 \)

where \( \Omega \ni x \mapsto 1(x) = 1 \).

We say that a generator s.t. \( B 1 = 0 \) is conservative.

3.7 Infinitesimal \( G \) Scale-Space

Combining our results about infinitesimal generators of a \( G \) scale-space, we can now state:

**Definition 3.7.1.** A \( g \) scale-space wedge, is a minimal Lie wedge of negative definite conservative operators \( h \ni w \mapsto B_w : L(\Sigma, \Sigma) \), that is a covariant tensor operator (3.10), with respect on the Lie algebra action \( g \ni v \mapsto A_v : L(\Sigma, \Sigma) \) and the Lie algebra and Lie wedge representation \( w \mapsto (v \mapsto C(v)(w)) \).

Summarizing the discussion above:

**Theorem 3.7.1.** A \( G \) scale-space is generated (using (3.12)) by its corresponding \( g \) scale-space wedge.

Now we have the tools we need for being able to study infinitesimal \( G \) scale-spaces. We will apply these tools for Euclidean similarity spaces, affine spaces and Galilean spaces with scaling in time and space.
Chapter 4

Spatio-Temporal Scale-Spaces

4.1 Introduction

In this chapter we will apply the theoretical framework from Chapter 3 on a number of geometries that are relevant for visual applications. We start with the affine line in Section 4.2 and generalize earlier results by showing that there is a two parameter family of scale-spaces on the affine line that are generated by Feller fractional derivatives. The corresponding convolution kernels are stable density functions and are well known from probability theory. This family of scale-spaces contains previously studied reflection symmetric scale-spaces and time causal scale-spaces as sub cases.

We continue by applying the framework on Euclidean similarity (Section 4.3) reproducing the results from Pauwels, VanGool, Fiddelaers & Moons (1995).

Next (in Section 4.4) we start by defining a spatio-temporal scale-space suitable for image motion measurements in the seemingly most natural way: It should be a semi-group on space-time, time causal and fulfill Galilean covariance. We require Galilean covariance as for a moving observer no particular motion or lack of motion should be treated in a special way. It is relative motion that matters. This approach however doesn’t lead anywhere as we show that among the possible Galilean covariant, semi-groups on space-time, there are no time causal ones.

After the failure of the “obvious” approach to time causal spatio-temporal scale-space, we perform a deeper analysis of the nature of temporal measurement in Section 4.5. And observe that in addition to that a real time system cannot be dependent on future information it cannot be directly dependent on past input either, only its representation of past input, its memory (Fagerström 2003). Based on this insight, we instead define a temporally causal spatio-temporal scale space as a semi-groups in space and memory, and derive a family of scale-spaces fulfilling this (Section 4.6). If one requires the generator of the scale-space to be local, there is a unique scale-space that is generated by the heat equation.
4.2 The Affine Line

A one dimensional scale space measurement is invariant w.r.t. translation and covariant w.r.t. scaling. The affine line has the infinitesimal generators,
\[ g_l(1) = t(1) \cup \{x \partial_x\}, \quad (4.1) \]
commutator \([\partial_x, x \partial_x] = \partial_x\), and the symbol for scaling is, \(i^{-1}x\xi\).

**Lemma 4.2.1.** Translation invariant symbols \(b : \mathbb{R} \rightarrow \mathbb{C}\) covariant with \(g_l(1)\) are on the form,
\[ b(\xi) = k\xi^\alpha \]
for any \(k \in \mathbb{C}\) and \(\alpha \in \mathbb{R}\).

**Proof.** From translation invariance, the covariance tensor must obviously be trivial for the translation generator, for scaling the simplest non-trivial representation is \(C(x \partial_x) = \alpha\), for any \(\alpha \in \mathbb{R}\). Combining this with, (3.10) we have,
\[ [i^{-1}x\xi, b(\xi)] = \alpha b(\xi), \]
for the symbol \(b(\xi)\). Using (3.18), we get
\[ \xi b'(\xi) = \alpha b(\xi) \]
which has solutions on the form given in the lemma.

By parameterizing \(k\) as \(k = ce^{i\theta \pi/2}\), where \(c, \theta \in \mathbb{R}\) and disregarding the, in this context, uninteresting scaling parameter \(c\), we can describe the above result in terms of fractional derivatives:

**Definition 4.2.1.** The **Feller fractional derivative** (Samko, Kilbas & Marichev 1992) is defined as,
\[ D_\theta^\alpha (\xi) = -e^{i\theta \pi/2} \xi^\alpha = -|\xi|^\alpha e^{i\text{sign}(\xi) \theta \pi/2}, \quad (4.2) \]
where \(0 < \alpha \leq 2\), \(\alpha \neq 1\) and \(|\theta| \leq \alpha\) for \(0 < \alpha < 1\) and \(|\theta| \leq 2 - \alpha\) for \(1 < \alpha \leq 2\).

It can be shown that the Feller derivative is negative definite and conservative for the values of \(\alpha, \delta\) given in the definition. The 3 parameter family of functions generated of the Feller derivative are called **stable densities** (Feller 1966) and appear in generalizations of the central limit theorem. The parameter \(\alpha\) is called the order and describe the scaling properties, \(\theta\) describe the amount of asymmetry. Symmetric stable densities, \(\theta = 0\) was the result of the scale space axiomatization in (Pauwels et al. 1995). The maximally asymmetric, extremal stable density functions can be shown to be one sided for \(0 < \alpha < 1\) and \(\theta = \pm \alpha\), these where the result of an axiomatization of scale spaces with temporal causality in (Fagerström 2003). The stable density for \(\alpha = 2, \theta = 0\) is the normal distribution, for \(\alpha = 1, \theta = 0\)
it is the Cauchy distribution and for $\alpha = \theta = 1/2$ is the solution of the signaling equation, i.e. diffusion on the half line with the signal as input at the end.

In some cases it is useful to express the Feller derivative as a linear combination of a symmetric and an antisymmetric operator,

$$ D_\theta^\alpha (\xi) = -\cos(\theta \pi / 2) |\xi|^\alpha - \sin(\theta \pi / 2) i \text{sign}(\xi) |\xi|^\alpha. $$

**Definition 4.2.2.**

$$ D^\alpha \sim |\xi|^\alpha $$

is called the *Riesz fractional derivative*. For $\xi \in \mathbb{R}^d$ we will also use the notation

$$ -(\Delta)^{\alpha/2} = D^\alpha,$$

for emphasizing that the derivative can be seen as a generalization of the Laplacian.

Another useful decomposition is as a linear combination of the two one sided operators,

$$ D_\theta^\alpha (\xi) = -c_+ (\alpha, \theta) D_+^\alpha (\xi) - c_- (\alpha, \theta) D_-^\alpha (\xi). $$

**Definition 4.2.3.** $D_+^\alpha$ ($D_-^\alpha$) is called the left (right) sided Riemann Liouville fractional derivative and is given by,

$$ D_\pm^\alpha (\xi) = (\mp i \xi)^\alpha = |\xi|^\alpha e^{\mp i \text{sign}(\xi) \alpha \pi / 2}, \quad c_\pm (\alpha, \theta) = \frac{\sin[(\alpha \mp \theta) \pi / 2]}{\sin(\alpha \pi)}. \quad (4.3) $$

We summarize these results with:

**Theorem 4.2.2.** A $\mathfrak{gl}(1)$ scale-space wedge is generated by a Feller fractional derivative $D_\theta^\alpha$ with parameters according to the definition. The subspace of reflection symmetric wedges are generated by Riesz fractional derivatives $D^\alpha$ with $0 < \alpha \leq 2$ and temporally causal wedges by left sided Riemann Liouville fractional derivatives $D_+^\alpha$ with $0 < \alpha < 1$.

### 4.3 Similarity

The Euclidean similarity group on $\mathbb{R}^2$ consists of translation in the plane rotation and scaling. Scaling is generated by $s(2) = \{ s = x_1 \partial_1 + x_2 \partial_2 \}$ and rotation by $so(2) = \{ r = x_2 \partial_1 - x_1 \partial_2 \}$, and the Euclidean similarity algebra by,

$$ \mathfrak{es}(2) = \mathfrak{t}(2) \cup \mathfrak{s}(2) \cup \mathfrak{so}(2), $$

where the non-zero commutators are $[\partial_j, s] = \partial_j$, $[\partial_1, r] = -\partial_2$ and $[\partial_2, r] = \partial_1$.

**Theorem 4.3.1.** A $\mathfrak{es}(2)$ scale-space wedge is generated for any $0 < \alpha \leq 2$ by the Riesz fractional derivative $-(\Delta)^{\alpha/2}$. 

Proof. Starting from (3.21) and requiring infinitesimal rotational invariance,
\[ \mathbf{0} = \left[ i^{-1}(x_2 \xi_2 - x_1 \xi_1), b(\xi_1, \xi_2) \right] = (\xi_2 \partial_{\xi_1} - \xi_1 \partial_{\xi_2}) b(\xi_1, \xi_2) \] (4.5)
we get that, \( b(\xi_1, \xi_2) = b(\xi_1^2 + \xi_2^2) \). Combining that with scaling covariance,
\[ \left[ i^{-1}(x_1 \xi_1 + x_2 \xi_2), b(\xi_1^2 + \xi_2^2) \right] = \frac{\alpha}{2} b(\xi_1^2 + \xi_2^2), \] (4.6)
setting \( \psi = \xi_1^2 + \xi_2^2 \), that is equal to, \( \psi b'(\psi) = \frac{\alpha}{2} b(\psi) \) considering the symmetry of the situation only symmetric solutions in term of the Feller derivative are allowed, \( b(\xi_1, \xi_2) = |\xi_1^2 + \xi_2^2|^{\alpha/2} \), with 0 < \( \alpha \leq 2 \) as for the affine line.

4.4 Galilean Similarity

The 1+1 dimensional Galilean similarity group, i.e. translation invariance in space and time, separate scaling in space and time and Galilean boost in space time, have the following set of infinitesimal generators,
\[ \gamma s(2) = t(2) \cup s(1) \oplus s(1) \cup \gamma (1), \] (4.7)
where \( \gamma (1) = \{ \gamma = x_0 \partial_1 \} \) is the Galilean boost that “skew” space-time and \( s(1) \oplus s(1) \) is a direct sum of the scaling generator in space and time respectively. The non-zero commutators are
\[ [\partial_j, x_j \partial_j] = \partial_j, \ [\partial_0, \gamma] = \partial_1, \ [x_0 \partial_0, \gamma] = \gamma, \ [x_1 \partial_1, \gamma] = -\gamma. \]

Theorem 4.4.1. A \( \gamma s(2) \) scale-space wedge is generated by \( \{ \partial_2, \partial_0 \partial_1, \partial_1^2 \} \).

Proof. Requiring our Lie wedge separately covariant w.r.t. scaling booth in space and time and using the results from scale space on the affine line we can see that any wedge must contain two generators on the form
\[ b_j(\xi_0, \xi_1) = k_j \xi_0^{\alpha_j}, \ j = 0, 1. \]
Applying the Galilean boost, which has the symbol
\[ \gamma = i^{-1} x_0 \xi_1 \]
on these the spatial generator disappears
\[ [\gamma, k_1 \xi_1^{\alpha_1} ] = 0, \]
while repeated application of the Galilean boost on the temporal generator gives,
\[ \text{ad}(i^{-1} x_0 \xi_1)^j (k_0 \xi_0^{\alpha_0}) = k_0 \prod_{j=0}^{i-1} (\alpha_0 - j) \xi_0^{\alpha_0 - j} \xi_1^j, \]
where
\[ \text{ad}(a)(b) = [a,b]. \]
To get a finite base of generators we must have \( \alpha_0 \in \mathbb{Z}_+ \), furthermore to generate a positive semi-group \( \alpha_0, \alpha_1 \leq 2 \). For \( \alpha_0 = 2 \) we get the set of symbols,
\[ \{\xi_0^2, \xi_0 \xi_1, \xi_1^4\}, \]
k_0 \in \mathbb{R} for \( \alpha = 2 \) (then only \( \theta = 0 \) is allowed). This set of generators are both closed under the Galilean similarity group and complete by choosing \( \alpha_1 = 2, k_1 = 1 \) for the spatial generator.

It should be noted that the generated scale-space is symmetric both in time and space and thus no time causal scale-spaces are given with this axiomatization. And as \( \gamma_s(n), n \geq 2 \) have \( \gamma_s(2) \) as a sub algebra, no time causal scale-spaces are possible for them either.

The 2 + 1 dimensional Galilean similarity group is generated by,
\[ \gamma_s(3) = cs(2) \oplus gl(1) \cup \gamma(2) \quad (4.8) \]
where \( \gamma(2) = \{x_0 \partial_1, x_0 \partial_2\} \).

With similar arguments as in the 1 + 1-dimensional case the lie wedge can be shown to have symbols on the form,
\[ \sigma(\xi - v \xi_0) \cdot (\xi - v \xi_0) + \tau \xi_0^2, \quad (4.9) \]
where \( \xi = (\xi_1, \xi_2), v = (v_1, v_2) \) is the velocity, \( \sigma \) spatial scale and \( \tau \) temporal scale, (Florack, ter Haar Romeny, Koenderink & Viergever 1992a).

Also in this case time causality is not possible.

4.5 Time Causal Scale-Spaces

This far we have seen that Galilean scale-spaces as we have defined them have kernels that are symmetric in the temporal direction. This means that both the past and the future signal is used for the measurement, this is a disappointment if we want to use such a scale space for real time measurements.

**Definition 4.5.1.** The history at time \( t \) of the spatio-temporal signal \( \mathbb{R} \times \mathbb{R}^d \ni (t,x) \mapsto f(t,x) \in \mathbb{R} \) is,
\[ f(t,s,x) = \begin{cases} f(t-s, x) & \text{if } s < t \\ 0 & \text{otherwise}. \end{cases} \]

A *time causal* measurement operator only depends on the history at time \( t \) for a measurement at \( t \).
For a real time system temporal causality is obviously necessary. But as discussed in (Fagerström 2003) defining temporal measurement in terms of a convolution or an evaluation equation on the history is to beg the question. It would require the measurement device to already have access to what it is supposed to measure.

This conceptual problem can be resolved by involving a memory of the history in the definition of temporal measurement. The measurement device should only have access to the current signal and its memory of previous measurements. As the memory is supposed to represent the history it is reasonable to make it as similar to the history as possible. This can be done by requiring the memory to be a half-space of the same dimensions as the history and subject to the same symmetry requirements.

We apply these considerations on the pure time causal scale-space (with no spatial dimensions).

**Definition 4.5.2.** A time causal scale-space on the affine line \( \mathbb{R} \times \mathbb{R}_+ \ni (t, \tau) \mapsto u(t, \tau) \), where \( t \) is the temporal coordinate and \( \tau \) the memory coordinate, is generated by the signaling problem,

\[
\begin{align*}
\partial_t u &= Bu \\
u(t, 0) &= f(t),
\end{align*}
\]

where the operator \( B \) is independent of time and \( f : \Sigma \) is the input signal. The measurement operator implicitly defined by \( u(t, \tau) = \Phi_\tau f(t) \) is a GL(1) covariant point measurement operator (but not necessarily a semigroup) and the infinitesimal generator \( B \) is a \( \mathfrak{gl}(1) \) scale-space wedge.

**Theorem 4.5.1.** A time causal scale-space on the affine line is generated by the right sided Riemann Liouville fractional derivative \( D^\alpha_{\tau^-} \) (on the memory domain), with \( 1 < \alpha \leq 2 \).

**Proof.** The form of the generators of the \( \mathfrak{gl}(1) \) scale-space wedge is given in Theorem 4.2.2, of the generators given there, only \( B = D^\alpha_{\tau^-}, \ 0 < \alpha \leq 2 \) are translation invariant on the right (positive) half line, as all other generators have support on the left half line as well (for non integer differentiation order).

In (Fagerström 2003) it is shown that (4.11), with \( B = D^\alpha_{\tau^-} \) is equivalent to

\[
\begin{align*}
\partial_\tau u &= -D^{1/\alpha}_{t^+} u \\
u(t, 0) &= f(t),
\end{align*}
\]

where \( 0 < 1/\alpha < 1 \) that is \( \alpha > 1 \) for \( \Phi_\tau \) to be a time causal scale space.

It is worth noticing that the only local generator for this family of scale spaces is \( D^2_{\tau^-} = \partial^2_\tau \), which means that in this case the scale space is generated by the heat equation, although with different boundary conditions compared to ordinary spatial scale space.
4.6 Time Causal Galilean Scale-Spaces

Now we extend these results to Galilean space-time. The time causal Galilean scale-space is as discussed above a scale space on space and memory rather than on space and time. Something new compared to the the previously discussed scale-spaces is that for the generator of the scale space on memory \( \partial_t = D^\alpha_{\tau} \), the symmetry Lie algebra will not only act on the generator but on the time derivative \( \partial_t \) as well.

Definition 4.6.1. Let
\[
\gamma_{d+1} = \mathfrak{c}(d) \oplus \mathfrak{gl}(1) \cup \gamma(d).
\]
The \((d+1)\)-dimensional time causal Galilean scale-space
\[
\mathbb{R}^+ \times \mathbb{R}^d \times \mathbb{R}^+ \times \mathbb{R}^d \ni (\sigma, v, \tau, t, x) \mapsto u(\sigma, v, \tau, t, x) \in \mathbb{R},
\]
where \( \sigma \) is spatial scale, \( v \) is velocity, \( \tau \) is memory (and temporal scale) is a \( \gamma_{d+1} \)-covariant, point measurement space in space-time \((t, x)\) and a \( \mathfrak{gl}(1) \)-wedge in memory \( \tau \).

Theorem 4.6.1. A \((d+1)\)-dimensional time causal Galilean scale-space (for \( d = 1, 2 \)) is generated by the evolution equation,
\[
\begin{cases}
\partial_t u = -v \cdot \nabla_x u + D^\alpha_{\tau} u \\
\partial_\sigma u = -(-\Delta_x)^{\alpha/2} u \\
u(0, 0, 0, t, x) = f(t, x),
\end{cases}
\tag{4.13}
\]
where \( 1 < \alpha_0 \leq 2, 0 < \alpha \leq 2, \) \( \sigma \) is the spatial scale direction, \( v \in \mathbb{R}^d \) the velocity vector, \( \partial_\sigma = \partial_t + v \cdot \nabla_x \) is the spatio-temporal direction, \( \nabla_x = (\partial_1, \ldots, \partial_d) \) is the spatial gradient and \( \Delta_x \) is the spatial Laplacian.

The equation,
\[
\begin{cases}
\partial_t u = -v \cdot \nabla_x u + \partial_\sigma^2 u \\
\partial_\sigma u = \Delta_x u \\
u(0, 0, 0, t, x) = f(t, x),
\end{cases}
\tag{4.14}
\]
is unique in the family of evolution equations as it is the only that has local generators.

Proof. First it is shown in Theorem 4.3.1 that the spatial part of the scale space is generated by \( \partial_\sigma u = -(-\Delta_x)^{\alpha/2} u, 0 < \alpha \leq 2 \) that besides being covariant with \( \mathfrak{c}(2) \) is invariant with respect to temporal translation and scaling and spatio-temporal Galilean boost. For the temporal part of the scale-space we know from Theorem 4.5.1 that the generator \( \partial_t u = D^\alpha_{\tau} u, 1 < \alpha \leq 2 \) is a \( \mathfrak{gl}(1) \)-wedge. As \( D^\alpha_{\tau} \) is independent of space and time it is obviously invariant with the Galilean similarity group. But \( \partial_t \) is not checking for the commutation relations with \( \gamma_{d+1} \) the non zero commutators are for scaling \([\partial_t, t\partial_t] = \partial_t\) and for the Galilean boosts \([\partial_\sigma, t\partial_j] = \partial_j, i = 1, 2\). Checking \( \partial_j \) for all commutators no further generators are
added, so \( \{ \partial_t, \partial_j \} \) is closed under \( \gamma \sigma(d + 1) \). As a result, linear combinations of \( \{ \partial_j \} \), that is \( v \cdot \nabla_x, v \in \mathbb{R}^d \) must be added to the temporal scale-space generator to make it closed under the required symmetries.

While there is no closed form for the general causal Galilean scale-space, it can be shown that (4.14) have the solution

\[
\begin{align*}
  u(\sigma, v, \tau, t, x) &= \phi(\sigma, v, \tau, \cdot, \cdot) * f(t, x), \quad (4.15) \\
  \phi(\sigma, v, \tau, t, x) &= \tau \exp\left(-\frac{x^2}{4t} - \frac{(x-tv)^2}{4\sigma}\right) \\
                          &= \sqrt{\frac{\pi}{4\pi t^{3/2}}} \frac{1}{(4\pi \sigma)^{d/2}}. \quad (4.16)
\end{align*}
\]

Figure 4.1: Causal Galilean scale-space: space vertically and time horizontally, three different velocity adaptations.

4.7 Discussion

The main result of this chapter is that we have shown (Theorem 4.6.1), that there is a reasonable spatio-temporal velocity adapted scale space for an active observer. It is booth time causal and recursively generated, in the sense that it only depends on current input and its memory, not the history of the input. The set of axioms is very close to modern scale-space axiomatizations (Weickert, Ishikawa & Imiya 1999), but with the main difference that we apply them on space and memory instead of space and time.

Comparing to earlier formulations of time causal Galilean scale spaces Lindeberg’s (Lindeberg 2002) is close in the sense that it is a recursive formulation, but the formulation is discrete so covariances are only approximate and it is much harder analyze the properties. The approach from Florack (Florack 1999) is based on a Gaussian Galilean scale-space that is made time causal by doing a logarithmic transformation of the time domain, it depends on the history of the signal and no recursive formulation has been suggested. Salden have proposed a time causal spatio-temporal scale-space where the diffusion equation is applied (separately on
the spatial and temporal domain) on the history of the signal so it is obviously de-
pendent on history rather than recursive. Although the original formulation is not
Galilean, it could easily be, by using a Galilean transformation on the generator.

Recently Lindeberg (2010) has derived a time causal Galilean scale-space based
on an axiomatization where non-enhancement of local extrema is used instead of
on scale covariance. His axiomatization leads to the same scale-space as the cur-
rent work. He also generalizes the theory by allowing affine geometry instead of
Euclidean similarity as spatial geometry.

For numerical implementation of the time causal Galilean scale-space the heat
equation scheme (4.14) is most attractive as the fractional derivatives are integral
operators and need to involve a much larger number of grid points for getting
satisfying precision. For the heat equation it should be noted that it consists of two
independent heat equations one in space and one in space time. The one in space
can be computed with an explicit scheme for the heat equation with sub sampling
as described in (Lindeberg & Bretzner 2003). For the spatio-temporal part the
scheme in (Fagerström 2003) can be used.
Chapter 5

Article Summaries

5.1 Article I: Scale-Space with Causal Time Direction

Aim and Background

The aim of this article is to develop a multi scale representation for temporal visual signals that respects the causality of the time dimension. I.e. only information from the past can be used for calculating the image representation on courser time scales. A further aim is to make the theory practically useful for a real time visual systems by developing numerical algorithms and a computational architecture for calculating multi scale time (and space) derivatives in an efficient way.

An earlier approach for a temporal scale-space theory was developed by Koen-derink (1988). In this approach, the time axis was logarithmically transformed, so that the present moment was mapped to the infinity. On this transformed time axis, ordinary spatial Gaussian scale-space was applied. In this article, the time causal scale-space is axiomatically constructed on the untransformed time-axis.

In the scale-space literature, several different set of axioms have been used that, perhaps surprisingly, lead to Gaussian kernels as the unique choice. The current article starts from the scale-space axioms from (Lindeberg 1990, Lindeberg 1993):

- non creation of local extrema (zero-crossings) with increasing scale
- linearity
- shift covariance
- continuous semi-group structure
- reflectional symmetry.

With this set of axioms, Gaussian convolution and a rescaling of the modified Bessel function of integer order are the unique solutions. The second result gives a discrete analogue to the Gaussian scale-space.
CHAPTER 5. ARTICLE SUMMARIES

Approach and Classification

To create a time-causal temporal scale-space, we start from the above set of axioms (Lindeberg 1990, Lindeberg 1993), ignore continuous semi-group structure and replace reflectional symmetry (a spatial property) with: time causality (i.e. the kernel must be zero for future signals).

Based on classification results from Schoenberg (1953) there are 4 primitive classes of kernels that satisfy the current temporal scale-space axioms, one of them is continuous:

- truncated exponential functions

and three of the classes are discrete kernels:

- normalized Poisson kernel.
- two-point weighted average
- moving average or first order recursive filtering

We will look at the cases for continuous and discrete time separately.

Continuous Time

For continuous time the only possible kernels fulfilling the chosen set of axioms are truncated can be constructed by convolve a sequence of such kernels with possibly varying parameters. It can be shown that we cannot create a scale-space with a continuous scale by some limit process where we let the scale of the individual truncated exponential kernels in the convolution going towards zero (the result is a Dirac pulse). And as a result a continuous scale is not possible for continuous time, the scale parameter must be discrete. Furthermore there is no unique scale-space as different choices of the sequence of parameters for the convolved kernels, leads to different shape of the resulting kernel (see figure 5.1).

Figure 5.1: Cascaded truncated exponentials
5.1. ARTICLE I: SCALE-SPACE WITH CAUSAL TIME DIRECTION

Discrete Time

For the three classes of kernels that generate discrete temporal scale-spaces, the normalized Poisson kernels are unique in that they form a continuous convolution semi-group in the scale dimension. The semi-group is generated by a discrete first order temporal difference (figure 5.2).

![Figure 5.2: Poisson kernels](image)

The two remaining classes of temporal scale-space kernels; moving average and two-point weighted average are less interesting from an analytical point of view, as they generate temporal scale-spaces that are discrete in both time and scale. And in that there are multiple levels of freedom in the choice of parameters, so no unique scale-space is singled out.

On the other hand they have excellent computational properties as each kernel just requires a few arithmetical operations and a scale-space can be constructed by cascade coupling a sequence of the basic kernels. Also no additional time buffering is needed as the cascade coupled kernels already acts as time buffers.

Furthermore moving average and two point average are basic components of digital filter theory, so their properties are well known and there are well developed analytical tools for studying their properties.

Spatio-Temporal Scale-Space Architecture

For practical usage of the temporal scale-space the scale levels of the scale-space must be distributed in some way, temporal derivatives needs to be computed and the temporal scale-space needs to be combined with a spatial scale-space.

As scale is multiplicative, it is natural to distribute the scales in a geometric series, so that the quote of two consecutive scales is constant. Parameter values needed for creating a geometrically distributed scale-space from cascade coupled truncated exponentials or moving average filters are derived in the article.

It is also shown that temporal derivatives can be calculated from differences between temporal scales. This means that there is no need to store the temporal scale space from earlier moments in time for calculating temporal derivatives.

A simple spatio temporal scale-space can be created from a direct product of a spatial and a temporal scale-space. An efficient way of doing this is by calculating
the spatial scale-space in the finest temporal scale and then calculate a temporal scale-space for each spatial scale (figure 5.3).

5.2 Article II: Temporal Scale Spaces

Aim
While we succeeded with our aim to develop an axiomatic time causal scale-space theory in article I, the resulting set of scale spaces must either be discrete in time or scale and lack scale covariance. This makes them less convenient to use as a basic building block in a theory about temporal observation. In article II, the aim was to develop time causal scale-space that is continuous in both time and scale while still being derived from a reasonable set of axioms.

Approach and Result
As in article I we start from a set of axioms that generate a spatial scale space and replaces the axiom of reflectional (or rotational) symmetry with the axiom of temporal causality. But instead of using the axioms from Lindeberg (1990) we instead use the axioms from Pauwels et al. (1995). We require a causal scale-space kernel to have the properties of:

- continuity
- positivity
- unit area
- temporal causality
- dilation covariance
- convolution semigroup
The main difference between this set of axioms is that in article I is that we use scale covariance instead of non-creation of local extrema. It should be noted that the combination of unit area, positivity and convolution semi-group makes the resulting scale-spaces dissipative and that they removes structure with increasing scale. But this is a weaker requirement for non creation of structure, than non-creation of local extrema.

We show that there exists a unique one parameter family (0 < a < 1) of kernels fulfilling this set of axioms. The resulting causal scale-space kernels are well known in the field of probability theory as extremal stable density functions.

**Recursive Formulation**

We then continue to state that for a theory about temporal measurement it is not enough to require that the observer cannot access the future, also the observer does not have any direct access to the past. It only has an indirect access to the past through its memory of it.

We show that the family of scale spaces can be described in terms of integrodifferential evolution equations both on the temporal and the scale axis. The evolution equation on the scale axis with the input signal as boundary condition on the temporal axis is especially interesting as the the scale space is the only memory of the past signal in this case.

The evolution equations have fractional derivatives (a generalization of ordinary derivatives to non integer orders) as infinitesimal generators. For the parameter $a = 0.5$ the fractional derivative is an ordinary 2:nd derivative and the evolution equation becomes a diffusion equation, but on the scale and not on the signal as in spatial scale space. This evolution equation is known as the signaling equation.

An efficient numerical implementation of the signaling equation is developed. To reduce the number of grid points it is natural to use log spaced grid points in the scale direction, we derive a discrete second derivative for a grid with geometrical progression. The resulting numerical scheme can be implemented with 10 operations per scale and time step.

**5.3 Article III: Spatio-Temporal Scale-Spaces**

**Aim**

In article III the aim is to develop a time causal spatio-temporal scale-space theory suitable for image motion measurements. A simple time causal spatio-temporal scale-space can be constructed by just applying the cross product of ordinary Gaussian scale-space in the image plane and a time causal scale-space according to article II in time. But the resulting spatio-temporal scale-space favours non moving objects. For a moving observer, observing moving objects relative motion matters. Following earlier work (Florack, ter Haar Romeny, Koenderink & Viergever 1992b) we formalize this by requiring the scale-space to be co-variant with respect to
Galilean transformations i.e. no particular motion or lack of motion should be treated as special.

**Approach and Theory**

We start from a set of axioms that includes the ones from article II, but split them into three parts to make the different roles of the axioms clearer:

- **Point measurement operator**
  - linearity
  - gray level invariance (corresponds to unit volume for a kernel)
  - positivity
  - point (has the identity operator in the limit)

- **G-covariance w.r.t. Galilean similarity:**
  - translation in space and time
  - scaling in space and time
  - rotation in space
  - Galilean boost in space-time

- **Semi-group of G-covariant point measurements**

  We then develop infinitesimal criteria for the above axioms as it simplifies the development of the theory as it linearize the problem. Also it allows for more general boundary conditions (which we will need later).

  The infinitesimal object that correspond to the transformation group is a Lie algebra, and the infinitesimal object that correspond to the semi group is a Lie wedge of infinitesimal generators. The G-covariance criteria corresponds to being a covariant tensor operator, gray level invariance and positivity to that the operator is negative definite conservative operator.

  The infinitesimal scale-space theory is then applied on the affine line (translation and scaling) and it is shown that all one-dimensional scale-spaces are generated by Feller fractional derivatives with certain restrictions on the 2-parameters operator family. The subspace of reflection symmetric scale-spaces are generated by Riesz fractional derivatives and time causal scale-spaces by left sided Riemann Liouville fractional derivatives. This generalizes and give a common framework to the results in Pauwels at al (1995) and in article II.

  Applied on Galilean similarity, we show that the only possible scale space is generated by mixed second order derivatives. And these does not generate time causal scale-spaces. This far we have shown that there does not exists a time causal spatio-temporal scale-space that is generated on time and space.
5.4. **ARTICLE IV: GALILEAN DIFFERENTIAL GEOMETRY OF MOVING IMAGES**

**Time Recursive Approach**

In article II we showed that the time causal temporal scale-spaces have infinitesimal generators on both time and scale. We furthermore argued that it is preferable to generate them based on the scale axis, as the scale then also functions as a memory of the past signal.

We can use the same approach for spatio-temporal time causal scale-space and define it as generated on space and memory (i.e. scale) with the image sequence as boundary condition instead of generating it on space and time.

With this definition we show that there exists a family of scale-spaces and give the (fractional derivative) evolution equations that generate them. In this family there is only one member that has local generators (i.e. is a partial differential equation), in term of heat equations on space and memory respectively. For this scale-space we also derive a closed form.

For the locally generated Galilean scale-space efficient numerical implementations are possible as the temporal scale is the only needed memory of past signals. And the discrete differential operators for spatial-scale and temporal steps requires just a few grid points.

---

**5.4 Article IV: Galilean Differential Geometry of Moving Images**

**Aim**

In article IV the aim is to describe image motion directly in term of the local geometric structure of the moving (spatio-temporal) image. This differ from the common optical-flow based approach where image motion is described in terms of the motion of projected points from the surrounding scene. The projected motion of points can not in general (except for special cases like e.g. corners) be observed directly from the moving image but must instead be indirectly inferred from an area in the moving image.

**Approach**

As in article III, we argue that Galilean geometry is a natural background for studying image motion for a moving observer as it has relative motion and relative direction as invariants rather than absolute motion and direction. The moving (gray-level) image can be seen as a 2+1 dimensional spatio-temporal function and local structure means differential geometry. So from this perspective, studying local structure of image motion means studying 2+1 dimensional Galilean differential geometry of functions, and especially its invariants.

We derive Galilean differential invariants by using the Cartan theory of moving frames. This is done by connecting a Galilean frame (a coordinate system) to each point in the moving image in such a way that the frame is connected to the local
image structure (figure 5.4). Invariants are then found from studying how frames in the surrounding of a point are transformed relative to the frame attached to the point. All mixed derivatives in terms of the local frame are by construction invariants and are called Galilean differential invariants.

We focus on two ways to attach the Galilean frame to the local moving image structure that are especially interesting from a vision perspective. In the first, gradient gauge, we connect the frame to the isophote surface through a point. The isophote surface, is the surface swept out by the motion of a isophote curve in the image plane over time. The gradient gauge is well defined as long as the isophote surface is traversal to the image plane and has a non-diminishing curvature in the image plane. Invariants based on the gradient gauge are also invariant with respect to monotonic (global) gray level transformations (lightening variations). See figure 5.5 and 5.6 for examples.

In the second case, Hessian gauge, the frame is adapted to the local image structure in such a way that the mixed second derivatives along the frame disappears. This can be well defined even when the isophote surface is tangent to the image plane and can be used in points where structure disappear or is created, e.g. along motion boundaries. Invariants based on the Hessian gauge both with respect to global gray level transformation and local linear lightening gradients. See figure 5.7 and 5.8 for examples.

Results
We derive a full set of scalar invariants both in terms of the gradient gauge and the Hessian gauge. We furthermore show that both set of invariants can be described in terms of the Cauchy-Stokes decomposition theorem: curl, divergence and deformation together with an acceleration vector and some image plane invariants.
5.4. ARTICLE IV: GALILEAN DIFFERENTIAL GEOMETRY OF MOVING IMAGES

Figure 5.5: Example of pure tangent acceleration: \( a \neq 0, \delta = 0 \).

Figure 5.6: Example of pure tangent divergence: \( a = 0, \delta \neq 0 \).
Figure 5.7: Example of pure Hessian acceleration: $a \neq 0, \delta = 0$.

Figure 5.8: Example of pure Hessian divergence: $a = 0, \delta \neq 0$. 
Chapter 6

Conclusion and Open Issues

6.1 Conclusion

The main contribution from this thesis is a systematic space-time geometric approach to real time low level motion vision.

This is done in a number of steps:

- We argue for studying motion vision in terms of space-time geometry rather than in terms of a two (or a few) consecutive image frames. The use of Galilean Geometry and Galilean similarity geometry is motivated and relevant geometrical background is reviewed.

- To be able to measure the visual signal in a way that respects the geometry of the situation and the causal nature of time, we argue that a time causal Galilean spatio-temporal scale-space is needed. The scale-space axioms are chosen so that they generalize popular axiomatizations of spatial scale-space to spatio-temporal geometries.

- To be able to derive the scale-space a infinitesimal framework for scale-spaces that respects a more general class of Lie groups (compared to previous theory) is developed and applied.

- Perhaps surprisingly, we find that a with the chosen axiomatization, a time causal Galilean scale-space is not possible as an evolution process on space and time. But it is possible on space and memory. We argue that this actually is a more accurate and realistic model of motion vision.

- A set of low-level features are derived by classifying basic Galilean differential invariants.

While the derivation of the time causal Galilean spatio-temporal scale spaces required some exotic mathematics, the end result is as simple as one possibly could hope for and a natural extension of spatial scale-spaces. The unique infinitesimally

61
generated scale-space is an ordinary diffusion equation with drift on memory and
a diffusion equation on space. The drift is used for velocity adaption, the “velocity
adaption” part of Galilean geometry (the Galilean boost) and the temporal scale-
space acts as memory.

If we don’t restrict ourselves to infinitesimally generated scale-spaces we get
a family of scale-spaces that are generated by a family of fractional differential
evolution equations that generalize the ordinary diffusion equation. These kind of
evolution equations have become popular in recent research in e.g. financial and
physical modeling.

The Galilean differential invariants that are derived are equivalent with curl,
divergence, deformation and acceleration that normally are calculated in terms of
optical flow. But here they are calculated directly from the the spatio-temporal
image.

6.2 Open Issues

There are plenty of open issues, both theoretical and experimental. Except from
some simple illustrations of the numerical implementation of the scale-spaces, no
experimental evaluation is done at all.

Numerical Implementation

Numerical methods for diffusion equation with drift, which are needed for numerical
implementation of time causal Galilean scale-spaces, have been well known for a
long time. Still, work is needed for developing the details of how to actually use
it in this particular context. Also stable and efficient methods for calculating the
Galilean differential invariants are needed.

Sampling of the Scale-Space

The time causal Galilean scale-space has seven dimensions in the context of motion
vision. A naive implementation with constant grid distance along all dimensions
would produce an overwhelming amount of data. Fortunately, for the case of an
active visual observer the amount of data can be reduced to something more man-
ageable, by adapting the sampling to the actual information needs. Here the human
visual system can be a source of inspiration.

Along both spatial and temporal scale, it is more natural to have log spaced than
constant grid points. For spatial scale-space this is developed in pyramid schemes
and for temporal scale-space we developed numerical methods for log spaced grids.

In a visual detector, a certain minimal spatio-temporal volume is needed to
detect a photon. This means that a detector for a short time scale needs to integrate
over a larger spatial area and a detector for a fine grained spatial scale needs to
integrate over a longer extent of time. So there is no need for detectors that have
both the finest temporal and finest spatial scale.
For an active observer that follows the object of interest, it makes sense to have a foveated visual system, i.e. a visual system that has fine resolution in the central and coarser in the peripheral vision field. For an active foveated system, there will normally be small movements in the center and large movements in the periphery. Hence the Galilean boost (velocity adaption) can be sampled for small motions in the center of the visual field and for large movements in the periphery.

A foveated visual sensor that is sensitive for large motions in the periphery and for small motions in the center gives exactly the kind of input that is needed for building a control system for steering an active observer so that it can follow a moving object and stabilize it on its “retina”.

**Singularity Theory**

In spatial scale-space theory, qualitative geometry (singularity theory) of the scale-space in terms of scale-space events has been studied, both for the deep structure of the basic scale-space (Koenderink & van Doorn 1986) and for various features (Lindeberg 1999).

Given the four parameter measurement space from Galilean scale-space, what will the deep-structure look like, what structures will typically appear while varying the scale-parameters? To some extent our intuition can help us to generate hypotheses that then can be experimentally verified. The phenomena from spatial scale-space theory that fine-structure disappears with increasing scale will appear both for spatial and temporal scaling. Blobs will appear, disappear, melt together and split in two over time. For transparent motion of two patterns moving with different velocities, the first will appear when the velocity adapt to its velocity and the second will appear when the parameters match its velocity. In between the patterns will appear mixed to different degrees.

But even in the comparatively simpler case of spatial scale-space (two image dimensions and one scale dimension compared to three image dimensions and four parameter dimensions), intuitive expectations aren’t enough (Damon 1995). The only reasonable approach for understanding what can happen, is a systematic mathematical study of the structure of the measurement space.

The basic idea for studying generic properties of a mapping is to only consider such structures that survive small smooth perturbations. The intuition is that only such structures will ever be found in nature. A complete list of generic singularities for parametrized functions (with low parameter dimension), is Thom’s celebrated list of catastrophes (Gibson 1979).

This list cannot however be used directly for the case at hand. First, image space-time is, as already has been discussed, not any three dimensional space, it has Galilean structure. Some of the general perturbations used for Thom’s list mix time and space. As a consequence a more restricted set of perturbations that respects the Galilean structure, is needed. For a Galilean structure care must be taken about how the spatial plane cuts the singularity, which leads to an increased number of generic singularities. A list of space-time unfoldings (corresponding to
Thom’s list) respecting Galilean structure is given in (Wassermann 1975).

Second, the scale-space is not a general function space, it has the further restriction that must be the space of solutions to a partial differential equation. Methods for deriving generic properties for the heat equation are developed in (Damon 1995) and for more general classes of partial differential equations in (Damon 1997).

Work on the related problem of scale selection has been done on Lindeberg’s semi-discrete spatio-temporal scale-space in (Lindeberg 1997, Laptev & Lindeberg 2004).

Generalizing to Other Geometries

As a final note we mention that the $G$ scale-space framework that we developed for deriving Galilean scale-spaces and scale-spaces for its sub-geometries could possibly be used for deriving scale-spaces on other geometries.
Appendix A

Categories

We use concepts from diverse mathematical areas. Some kind of behaviour is common to all of these and it is convenient to have a vocabulary for these commonalities. We also need to compare mathematical objects from different areas. Category theory (Mac Lane 1971) is designed to deal with these tasks. We use some language and concepts from category theory to make our presentation more coherent.

Definition A.0.1. A category $\mathbf{C}$ consists of:

1. A set of objects, $X, Y, Z, \ldots \in \text{ob}(\mathbf{C})$.

2. For each pair of objects, $X, Y$, a set of morphisms, $f, g, \ldots \in \mathbf{C}(X, Y)$, written as $f : X \to Y$ or $X \xrightarrow{f} Y$, where $X$ is the domain, and $Y$ the codomain of $f$.

3. Composition: a set of maps:

\[
\mathbf{C}(X, Y) \times \mathbf{C}(Y, Z) \ni (f, g) \mapsto g \circ f \in \mathbf{C}(X, Z),
\]

for all $X, Y, Z \in \text{ob}(\mathbf{C})$.

Which fulfills the following axioms:

1. For each $X$, $\mathbf{C}(X, X)$ has an element $id_X$, which acts as a left and right identity, with respect to composition. For $f \in \mathbf{C}(X, Y)$:

\[
f \circ id_X = f = id_Y \circ f.
\]

We use the notation $id$, for identity if its domain is clear from the context.

2. Composition is associative. For $X \xrightarrow{f} Y \xrightarrow{g} Z \xrightarrow{h} W$:

\[
h \circ (g \circ f) = (h \circ g) \circ f,
\]

which makes parentheses unnecessary.
Examples A.0.1. Some common categories, (more are given in the rest of the appendix):

1. **Set**: Objects are sets and morphisms are maps.

2. **Vect**: Objects are vector spaces and morphisms are linear maps.

3. **Top**: Objects are topological spaces and morphisms are continuous maps.

Definitions A.0.2. A morphism $X \xrightarrow{f} Y$ is an isomorphism, if there exists a morphism $Y \xrightarrow{g} X$, such that $g \circ f = id_X$ and $f \circ g = id_Y$. A morphism $X \xrightarrow{f} X$ is called an endomorphism, and the set of endomorphisms on $X$ is denoted $\text{End}(X)$. Endomorphisms that are also isomorphisms are called automorphisms, $\text{Aut}(X)$. If $f \circ \sigma = id_X$ for $X \xrightarrow{\sigma} Y \xrightarrow{f} X$, $\sigma$ is said to be a section of $f$, (and $f$ a retraction of $\sigma$). If we need to emphasize that, an isomorphism, endomorphism, etc, belongs to a certain category $C$, we write $C$-isomorphism, $C$-endomorphism, etc.

These concepts are used everywhere in mathematics and have often category specific names e.g. Top-isomorphism is called homeomorphism.

It is useful to be able make comparisons between categories also. A functor is a morphism between categories.

Definitions A.0.3. For two categories $\mathbf{C}, \mathbf{D}$, a (covariant) functor $T : \mathbf{C} \to \mathbf{D}$, assigns to each object $X \in \mathbf{C}$ an object $TX \in \mathbf{D}$, and to each morphism $X \xrightarrow{f} Y \in \mathbf{C}$, a morphism $TX \xrightarrow{Tf} TY \in \mathbf{D}$, such that:

\[
T(id_X) = id_{TX}, \quad T(g \circ f) = Tg \circ Tf,
\]

whenever $g \circ f$, is defined in $\mathbf{C}$. A contravariant functor is defined as a covariant functor, but with the composition rule:

\[
T(g \circ f) = Tf \circ Tg.
\]
Appendix B

Calculus on Manifolds

In this appendix we gather some definitions and basic results (without proofs) from manifold theory. Our main aim is to make our chosen notation precise. For more complete accounts, see e.g. Bishop & Goldberg (1980), Hirsch (1976), Lang (1999), Spivak (1975) and Warner (1983).

B.1 Topological Prerequisites

See e.g. Armstrong (1983), Bishop & Goldberg (1980), and Simmons (1963).

Definition B.1.1. A topological space $X$ is a Hausdorff space if for all $x_1, x_2 \in X$, $x_1 \neq x_2$, there are neighborhoods $U_1, U_2$ respectively, such that $U_1 \cap U_2 = \emptyset$.

Compactness is a generalization of the concept of bounded metric spaces.

Definitions B.1.2. Let $X$ be a topological space then a class of open subsets $\{U_i\}$ of $X$ is an open cover of $X$ if $\bigcup U_i = X$. A subclass of an open cover of a topological space $X$ is said to be a subcover if it is also an open cover of $X$. A compact space is a topological space in which every open cover has a finite subcover.

Compactness is a topological property i.e. it is preserved by homeomorphisms, even stronger, compactness is preserved by onto continuous mappings. A compact subset of an Euclidean space is closed and bounded.

If a topological space is connected it consists of a single piece.

Definition B.1.3. A topological space $X$ is disconnected if it can be represented as the union, $X = U_1 \cup U_2$, of two disjoint open sets, $U_1 \cap U_2 = \emptyset$, it is connected if it is not disconnected.

Connectedness is a topological property.
B.2 Manifolds

A manifold is a locally Euclidean space, e.g. the surface of a sphere or the graph of a smooth function.

**Definitions B.2.1.** An \( n \)-dimensional topological manifold \( M \), is a Hausdorff space with countable basis that is locally homeomorphic with \( \mathbb{R}^n \). That is there is a set of homeomorphisms from open sets in \( M \) to open sets in \( \mathbb{R}^n \):

\[
M \supset U_i \xrightarrow{\varphi_i} V_i \subset \mathbb{R}^n,
\]

such that the set \( \{U_i\} \) is an open cover of \( M \). A pair \( (U_i, \varphi_i) \) is called a chart and the set \( \alpha = \{(U_i, \varphi_i)\} \) is called an atlas. An atlas is smooth if the map:

\[
\varphi_j \circ \varphi_i^{-1} : \varphi_i(U_i \cap U_j) \to \varphi_j(U_i \cap U_j),
\]

is \( C^\infty \) for each pair of charts in the atlas. It can be shown that for each atlas, there is an unique maximal atlas that contains the atlas. A topological manifold together with a maximal smooth atlas \( (M, \alpha) \) is called a smooth manifold.

**Definition B.2.2.** A map between two manifolds \( f : M \to N \) is smooth if for each chart \( (U, \varphi) \) for \( M \) and \( (V, \psi) \) for \( N \) such that \( f(U) \subset V \), the map

\[
\psi \circ f \circ \varphi^{-1} : \varphi(U) \to \psi(V)
\]

is \( C^\infty \). If a smooth map has an inverse that also is smooth it is called a diffeomorphism.

It can be shown that the composition of two smooth maps is still a smooth map. Smooth manifolds, together with smooth maps, forms a category, \textbf{Man}, where diffeomorphisms are \textbf{Man}-isomorphisms.

**Definition B.2.3.** Let \( m \in M \) be a point on a smooth manifold. Two smooth maps \( f_1 : U_1 \to N, f_2 : U_2 \to N \) defined on an environments \( U_1 \subset M, U_2 \subset M \) around \( m \) are said to have the same germ at \( m \) if they are equivalent on an environment \( U \) around \( m \). To have the same germ at \( m \), is an equivalence relation and the equivalence classes are called smooth germs at \( m \). Let \( \hat{F}_m \) denote the set of germs at \( m \) with codomain \( \mathbb{R} \). The operations of scalar multiplication, addition and multiplication on functions:

\[
(cf_1)(m) = c(f_1(m)), \quad cf_1 : U_1 \to N, c \in \mathbb{R}
\]

\[
(f_1 + f_2)(m) = f_1(m) + f_2(m), \quad f_1 + f_2 : U_1 \cap U_2 \to N
\]

\[
(f_1 \cdot f_2)(m) = f_1(m) \cdot f_2(m), \quad f_1 \cdot f_2 : U_1 \cap U_2 \to N,
\]

induces an algebra over \( \mathbb{R} \) on \( \hat{F}_m \).
Definition B.2.4. A subset $N$ of a an $m$-dimensional smooth manifold $(M, \alpha)$ is a $n$-dimensional smooth submanifold if there exists a smooth atlas $\alpha = \{(U_i, \varphi_i)\}$ for $M$ such that for each chart in the atlas
\[ \varphi_i(N \cap U_i) = V_i \times \{c\}, \]
where $V_i \subset \mathbb{R}^n$ and $c \in \mathbb{R}^{m-n}$.

Each map $\varphi_i$ induces a automorphism $\varphi'_i : N \cap U_i \to \mathbb{R}^n$,
and it can be shown that $\alpha' = \{N \cap U_i, \varphi'_i\}$ is an atlas for the smooth manifold $(N, \alpha')$.

We will often drop the prefix smooth for smooth manifolds, submanifolds, mappings etc.

B.3 Tangent Bundle

Derivatives for maps on manifolds does not stay on the manifold, there is need for a certain vector space, the tangent space for the derivatives of a map for each point on the manifold.

Definition B.3.1. A tangent vector $v$ at the point $m \in M$ is a linear derivation on the algebra $\tilde{F}_m$, i.e. for all $f_1, f_2 \in \tilde{F}_m$ and $c_1, c_2 \in \mathbb{R}$,
\[ v(c_1 f_1 + c_2 f_2) = c_1 v(f_1) + c_2 v(f_2) \]
\[ v(f_1 \cdot f_2) = f_1(m) \cdot v(f_2) + v(f_1) \cdot f_2(m). \]
The vector space of tangent vectors at $m$ is denoted the tangent space of $M$ at $m$, and we use the notation $T_m M$. The set of all tangent vectors on $M, TM = \{T_m \mid m \in M\}$ is called the tangent bundle of $M$.

It can be proven that each tangent space on a manifold has the same dimension as the manifold. There is a natural projection $\pi : TM \to M$ mapping elements in $T_m M$ on $m$. The tangent bundle $TM$ for an $n$-dimensional manifold is a $2n$-dimensional manifold.

Definition B.3.2. A smooth map $f : M \to N$ induces a linear map $d_m f : T_m M \to T_{f(m)} N$ at a point $m \in M$ of its domain, called the differential at $m$ defined as
\[ d_m f(v) = v(g \circ f), \]
where $v \in T_m M$ and $g : V \to \mathbb{R}$ is a smooth function that is defined in a neighborhood $V \subset N$ around $f(m)$. Differentials can be extended to tangent bundles $df : TM \to TN$ by
\[ df(m,v) = d_m f(v), \]
where $(m,v) \in TM$. 

It can be proven that for \( f : N \to O, g : M \to N \) that
\[
d(f \circ g) = df \circ dg.
\]

**Definition B.3.3.** A *vector field* is a smooth map \( X : M \to TM \) s.t. \( \pi \circ X = id \).

**Definition B.3.4.** Given two vector fields \( X, Y \) on \( M \), a new vector field \([X,Y]\) on \( M \), called the *bracket product* of \( X \) and \( Y \), be defined s.t. for each open set \( U \subset M \) and function \( f : U \to \mathbb{R} \),
\[
[X,Y](f) = X(Y(f)) - Y(X(f)).
\]

**B.4 Bundles**

Tangent bundles are examples of the more general concept of attaching some kind of structure at each point of a space with a certain structure in such a way that the new object gets the same kind of structure. Such object is called a *bundle*, (see e.g. Husemoller (1966), Lang (1999), Hirsch (1976) and Darling (1994)).

**Fiber bundles**

**Definition B.4.1.** A *fiber bundle* \((E, \pi, B, F)\) consists of topological spaces: \( E \), called *total space*, \( B \), called *base space*, \( F \), called *abstract fiber* and a continuous mapping \( E \xrightarrow{\pi} B \), called the *projection*. It is furthermore required that, for each neighborhood \( U \subset B \), \( \pi^{-1}(U) \) is homeomorphic to \( U \times F \) via a homeomorphism \( \varphi \), such that the following diagram commutes, i.e. each fiber is preserved.

\[
\begin{array}{ccc}
E \supset \pi^{-1}(U) & \xrightarrow{\varphi} & U \times F \\
\downarrow{\pi} & \swarrow{\text{proj1}} \\
U & & \\
\end{array}
\]

The pair \((U, \varphi)\) is called a *fiber chart* of \( E \). A collection of fiber charts \( \{(U_i, \varphi_i)\} \), such that \( \{U_i\} \) is an open cover of \( B \), is called a *fiber bundle atlas*. If each coordinate change:
\[
\varphi_j \circ \varphi_i^{-1} : (U_i \cap U_j) \times F \to (U_i \cap U_j) \times F,
\]
is smooth, the spaces \( E, B, F \) are manifolds, and the projection \( \pi \) is a smooth map, the fiber bundle is said to be a *smooth bundle*.

A coordinate change on a smooth bundle has the form
\[
\varphi_j \circ \varphi_i^{-1}(x,f) = (x, \varphi_{ji}(x)f),
\]
where \( \varphi_{ji} : U_i \cap U_j \to \text{Diff}(F) \) is a smooth map called a *transition function*. A bundle \((E, \pi, B, F)\), is often denoted \( F \to E \xrightarrow{\pi} B \) or \( E \xrightarrow{\pi} B \).
Example B.4.1. A simple example of a fibre bundle is the product bundle, \((B \times F, \text{proj}_1, B, F)\).

Definition B.4.2. A bundle morphism \((u, f) : (E, \pi, B, F) \rightarrow (E', \pi', B', F')\), between two fiber bundles, is a pair of two continuous maps \(E \xrightarrow{u} E'\) and \(B \xrightarrow{f} B'\), such that the following diagram commutes.

\[
\begin{array}{ccc}
E & \xrightarrow{u} & E' \\
\downarrow{\pi} & & \downarrow{\pi'} \\
B & \xrightarrow{f} & B'
\end{array}
\]

A smooth bundle morphism is a bundle morphism between two smooth bundles, such that the pair of maps \(u, f\) are smooth.

Vector bundles

A vector bundle is a fiber bundle with vector space structure on the fibers.

Definition B.4.3. An \(n\)-dimensional vector bundle \((E, \pi, B, \mathbb{R}^n)\), is a smooth bundle, where the fibers are \(n\)-dimensional vector spaces and the transition functions act with vector-space isomorphisms on the fibers, \(\varphi_{ji} : U_i \cap U_j \rightarrow \text{GL}(n)\).

Definition B.4.4. A vector bundle morphism is a smooth bundle morphism, \((u, f) : (E, \pi, B, \mathbb{R}^n) \rightarrow (E', \pi', B', \mathbb{R}^m)\), such that the restriction of \(u, u : \pi^{-1}(x) \rightarrow \pi'^{-1}(f(x))\), is linear for each \(x \in B\).

Vector bundles together with vector bundle morphisms, form the category \(\text{VB}\). We can also form the sub-categories \(\text{VB}_B\), by choosing a certain base space \(B\), and \(\text{VB}^n\), where the objects are \(n\)-dimensional vector bundles, and \(\text{VB}^n_B\), by fixating both dimensionality and base space.

B.5 Embeddings and Immersions

Definitions B.5.1. Let \(f : M \rightarrow N\) be a smooth map. Then \(f\) is an immersion if \(d_m f : T_m M \rightarrow T_{f(m)} N\) is an injection at each point \(m \in M\), \(f\) is a submersion if \(d_m f\) is a surjection at each point \(m \in M\), and \(f\) is an embedding if it is an injective immersion which is also a homeomorphism into its domain \(f(M)\).

B.6 Foliations

A foliation is a decomposition of an \(n\)-dimensional manifold in \(p\)-dimensional submanifolds, denoted leaves, that lie side by side, see e.g. Sharpe (1997).
Definitions B.6.1. An $q$-codimensional foliated atlas on an $n$-dimensional manifold, is an atlas $\alpha = \{(U_i, \varphi_i)\}$, where

$$\varphi_i : U_i \rightarrow \mathbb{R}^{n-q} \times \mathbb{R}^q,$$

such that coordinate changes $\Phi = \varphi_{ji}$ have the form,

$$(x, y) \mapsto (\Phi_1(x, y), \Phi_2(y)).$$

An $n$-dimensional manifold with an $q$-codimensional foliated atlas is called an $q$-codimensional foliated manifold.
Appendix C

Distributions

Distributions are defined in terms of test function spaces. There are many possible choices of test function spaces, or measurement devices in our application and we will spend a considerable effort to find a suitable class. As a starting point for our discussion however, we will list a number of less specialized test function classes that are commonly used in distribution theory.

Definitions C.0.2. Common test function spaces:

1. The space of smooth functions $C^\infty(\mathbb{R}^n)$, sometimes denoted $\mathcal{E}(\mathbb{R}^n)$.

2. The space of smooth functions with compact support $C^\infty_c(\mathbb{R}^n)$, sometimes denoted $\mathcal{D}(M)$.

3. The Schwartz class $\mathcal{S}$ or $\mathcal{S}(\mathbb{R}^n)$ (also called testing functions of rapid descent) iff $\sup |x^\beta \partial^\alpha \phi(x)| < \infty$ for all multi-indices $\alpha, \beta$. This means that the derivatives of the test functions decrease faster than the inverse of any polynomial.

4. The Sobolev spaces are both differentiable and integrable:

\[
W^k_p(\mathbb{R}^n) = \{ \phi : \partial^\alpha \phi \in L_p, |\alpha| \leq p, 1 < p < \infty, k \in \mathbb{N}_0 \},
\]

where

\[
L_p(\mathbb{R}^n) = \{ \phi : \left( \int |\phi(x)|^p \, dx \right)^{1/p} < \infty \}.
\]

We start by defining the most general form of distribution and list some of its properties (see e.g. Hörmander (1990) for details about distribution theory).

Definition C.0.3. A distribution $u$ over $X \subset \mathbb{R}^n$, is a linear form $C^\infty_c(X) \to C$, s.t. for each compact subset $K \subset X$ there exists $C$ and $k$ s.t.

\[
|u(\phi)| \leq C \sum_{|\alpha| \leq k} |\partial^\alpha \phi|, \quad \phi \in C^\infty_c(K).
\]

73
APPENDIX C. DISTRIBUTIONS

The set of all distributions in \( X \) is denoted \( \mathcal{D}'(\mathbb{R}^n) \), we will write \( \mathcal{D}' \) if the domain is obvious from the context. If the same \( k \) can be used for all \( K \), the distribution is said to be of order \( k \), and is denoted \( \mathcal{D}^k \). We will use the notions \( u(\phi) \) and \( \langle u, \phi \rangle \) interchangeably.

**Examples C.0.1.**

1. Each continuous function \( u_1 \in C^\infty(\mathbb{R}^n) \) can be identified with a distribution \( u \in \mathcal{D}'(\mathbb{R}^n) \), by,

\[
u(\phi) = \int u_1(x)\phi(x) \, dx, \quad \phi \in C^\infty(\mathbb{R}^n).
\]

Distributions that can be defined from functions in this way, are called regular distributions, (we will use the same symbol for both the regular distribution and the function that generates it, the meaning should be clear from the context).

2. There are also non-regular distributions, for example, the Dirac-"function" can be defined as the distribution,

\[
\delta(\phi) = \phi(0).
\]

3. A locally finite Borel measure \( \mu \) defined on \( X \subset \mathbb{R}^n \) determines a distribution by:

\[
u(\phi) = \int \phi \, d\mu, \quad \phi \in C^\infty_c(X).
\]

Conversely each zero order distribution is on that form by Riesz representation theorem.

4. Differential operators of order \( \leq k \) are examples of non-regular distributions of order \( k \).

If the distribution \( u \) represents the signal “out there”, \( \text{im } u = u(\Delta) \) represents the aspects of the world that our observer can perceive, given its measurement apparatus, \( \Delta \). The kernel, \( \ker u \), represents the observers ignorance, the aspects of the world that it can not discriminate between. The signal space \( \Sigma \) can be defined as the dual space of the space of test functions, \( \Sigma = \Delta' \).

A space of distributions can be made into a vector space by the obvious definitions:

**Definition C.0.4.** For \( u_1, u_2 \in \mathcal{D}' \), \( \phi \in C^\infty_c \), \( k \in \mathbb{R} \),

\[
(u_1 + u_2)(\phi) = u_1(\phi) + u_2(\phi)
\]

\[
(ku)(\phi) = k(u(\phi)).
\]

Derivation of distributions is defined by:
**Definition C.0.5.** For $u \in \mathcal{D}', \phi \in C_c^\infty$

$$(\partial^\alpha u)(\phi) = u((-1)^{|\alpha|}\partial^\alpha \phi).$$

(C.1)

with standard multi-index notation.

With this definition, we obviously don’t have to worry about the differentiability of the signals, it suffices to have differentiable test functions. The definition can be motivated by doing integration by part on regular distributions,

$$(\partial_{x_i} u)(\phi) = \int (\partial_{x_i} u)(x) \phi(x) \, dx = -\int u(x) \partial_{x_i} \phi(x) \, dx = -u(\partial_{x_i} \phi).$$

Where we use the fact that $\phi(x) \to 0$ when $|x| \to \infty$, for all $\phi \in C_c^\infty$.

The use of $C_c^\infty$ as test functions is to restrictive for our applications: the only analytic function with compact support is 0, so requiring compact support rules out e.g. Gaussians. Larger test function spaces results in smaller distribution spaces. We will, during our development of geometrical measurement, try to develop as narrow test function spaces as possible, fulfilling our requirements.

**Definitions C.0.6.** Continuous forms on $C^\infty(\mathbb{R}^n)$ are called *distributions with compact support*, $\mathcal{E}'$. Continuous forms on $\mathcal{S}(\mathbb{R}^n)$ are called *tempered distributions*, $\mathcal{S}'(\mathbb{R}^n)$.

We have that:

$$\mathcal{E} \supset \mathcal{S} \supset \mathcal{D}$$

and

$$\mathcal{E}' \subset \mathcal{S}' \subset \mathcal{D}'.$$
Appendix D

Lie Groups

See e.g. Warner (1983), Sagle & Walde (1973) and Onishchik (1993).

D.1 Abstract Groups

Definition D.1.1. An (abstract) group \( \{ G, \cdot \} \) is a set \( G \) including a special element \( \text{id} \) denoted identity, together with a binary associative operator \( \cdot \), denoted group multiplication, such that for all \( g \in G \),

\[
g \cdot \text{id} = g, \quad \text{id} \cdot g = g,
\]
and there must also be an unique inverse of each element in the group,

\[
g \cdot g^{-1} = \text{id}, \quad g^{-1} \cdot g = \text{id}.
\]

when the group multiplication operator is clear from the context, we will denote a group by its underlying set, and say e.g. the group \( G \).

Definition D.1.2. The group \( \{ H, \cdot \} \) is subgroup of the group \( \{ G, \cdot \} \) if \( H \subset G \) and the set \( H \) is closed under the group operation.

Definition D.1.3. The mapping \( F : G \to H \) is a group homomorphism from \( \{ G, \cdot \} \) to \( \{ H, \circ \} \) if

\[
F(g_1 \cdot g_2) = F(g_1) \circ F(g_2), \quad F(\text{id}_G) = \text{id}_H.
\]

The category \textbf{Grp} consists of groups together with group homomorphisms.

Definition D.1.4. The group \( G \) acts on the set \( X \) if there is a group homomorphism \( G \to \text{Aut}(X) \), \( g \mapsto \sigma(g) : \text{Aut}(X) \) i.e. for \( g_1, g_2, g \in G \)

\[
\sigma(g_1 \cdot g_2) = \sigma(g_1) \circ \sigma(g_2), \quad \sigma(\text{id}_G) = \text{id}_{\text{Aut}(X)}.
\]

\( \sigma(G) \subset \text{Aut}(X) \) is a transformation group. We will use the notation \( g \cdot x = \sigma(g)(x) \), and \( g \cdot f(x) = f(g \cdot x) \), for \( g \in G, f \in X \to Y, x \in X \).
Definitions D.1.5. Let $H$ be a subgroup of $G$. $H$ is a normal subgroup of $G$ if $gHg^{-1} = H$ for all $g \in G$. The quotient space $G/H$ is defined as

$$G/H = \{gH \mid g \in H\},$$

where $gH = \{gh \mid h \in H\}$ is called a left coset. A quotient space $G/H$ where $H$ is a normal subgroup of $G$ can be given a group structure by

$$(g_1 g_2)H = (g_1 H)(g_2 H),$$

and is called a quotient group.

D.2 Lie Groups

Definitions D.2.1. A Lie group is a group that which also is a smooth manifold and s.t. its group multiplication and inversion are smooth maps. A local Lie group, is a Lie group that only need to be defined in a neighborhood around the identity element. A Lie group which underlying manifold is compact, is called a compact Lie group.

Definition D.2.2. A subgroup $H \subset G$ of a Lie group $G$ is a (closed) Lie subgroup if the inclusion map $i : H \to G$ is a smooth embedding.

Let $G$ be a Lie group and $H \subset G$ be a normal Lie subgroup, then it can be shown that the quotient group $G/H$ is a Lie group.

Definitions D.2.3. A Lie group homomorphism is a smooth group homomorphism between two Lie groups. A Lie group homomorphism that is also a diffeomorphism is called an Lie group isomorphism.

Lie groups together with Lie group homomorphisms, form a category.

Definition D.2.4. A Lie group homomorphism $\sigma : \mathbb{R} \to G$ is called a one-parameter subgroup of $G$.

D.3 Lie Algebra

Definition D.3.1. A Lie algebra $\mathfrak{g}$ over $\mathbb{R}$ is a real vector space $\mathfrak{g}$ together with a bilinear operator $\mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$, $(x, y) \mapsto [x, y]$ (called the bracket) such that for all $x, y, z \in \mathfrak{g}$,

$$[x, y] = -[y, x] \quad \text{(anti-commutativity)}$$

$$[[x, y], z] + [[y, z], x] + [[z, x], y] = 0. \quad \text{(Jacobi identity)}$$
Definition D.3.2. A vector space homomorphism between two Lie algebras $\phi : g \to h$, is called a Lie algebra homomorphism if for all $x, y \in g$

$$\phi([x, y]) = [\phi(x), \phi(y)].$$

If $\phi$ furthermore is a vector space isomorphism it is a Lie algebra isomorphism.

Lie algebras together with Lie algebra homomorphisms, form a category.

Definition D.3.3. A vector field $X : G \to TG$ on the Lie group $G$ is left invariant if for all $g \in G$

$$dl_g \circ X = X,$$

where $l_g : \text{Diff}(G), G \ni x \mapsto gx$, is a left translation w.r.t. $g$. The set of left invariant vector fields on a Lie group $G$ is denoted by the corresponding lower case German character $\mathfrak{g}$ or by $\mathcal{L}(G)$.

Proposition D.3.1. Let $\mathfrak{g}$ be the set of left invariant vector fields on the Lie group $G$.

1. A left invariant vector field $X \in \mathfrak{g}$ is smooth.

2. The Lie bracket $[X, Y], X, Y \in \mathfrak{g}$ of two left invariant vector fields is a left invariant vector field.

3. $\mathfrak{g}$ together with Lie bracket forms a Lie algebra.

4. The map $\phi : \mathfrak{g} \to T_{id}G, X \mapsto X(id)$ is a vector space isomorphism and as a consequence $\mathfrak{g}$ and $G$ has the same dimension.

$T_{id}G$ can be extended to a Lie algebra by defining the the bracket operator $[\cdot, \cdot] : T_{id}G \times T_{id}G \to T_{id}G$ as

$$[x, y] = [\phi^{-1}(x), \phi^{-1}(y)](id).$$

Proposition D.3.2. The map $\phi : \mathfrak{g} \to T_{id}G$ defined above is a Lie algebra isomorphism.

1. Any Lie group $G$ possesses a (maximal?) connected solvable normal subgroup $N \subset G$ and the quotient group $G/N$ is semi simple.

2. Any $n$-dimmensional connected commutative real Lie group is isomorphic to $T^k \times \mathbb{R}^{n-k}$.

3. $G' = (G, G), (x, y) \mapsto xy^{-1}y^{-1}$ is called the commutator subgroup of $G$, it is normal and it is the smallest subgroup of $G$ s.t. its quotient, $G/G'$ is commutative.

4. An iterative commutator group $G^{(n)}$ is defined as $G^{(0)} = G, G^{(k)} = (G^{(k-1)})'$. 
5. A group is *solvable* if \( \exists n : G^{(n)} = \{ \text{id} \} \).

6. Any subgroup and quotient group of a solvable group is solvable.

7. If \( N \) and \( G/N \) are solvable, so is \( G \).

8. Any non-trivial simply connected Lie group decomposes in a semidirect product \( N \rtimes K \) where \( K \) is isomorphic to \( \mathbb{R} \) or \( \mathbb{C} \) and \( N \) is a normal subgroup.

9. There is a largest connected solvable normal subgroup, denoted \( \text{Rad}(G) \).

10. A Lie group is *semi simple* if \( \text{Rad}(G) = \{ \text{id} \} \), \( \forall G, G/\text{Rad}(G) \) is semi simple.

11. \( K \) is connected compact then \( K = ZK' \) where \( Z = \text{Rad}(K) \) is a compact torus.
Bibliography


