Invariance with Optic Flow

by

Yana Momchilova Mileva

Thesis handed in on: 31 May 2007

Supervisor:

Prof. Dr. Joachim Weickert

Advisor:

Dr. Andrés Bruhn

Reviewers:

Prof. Dr. Joachim Weickert

Prof. Dr. Reinhard Wilhelm

Mathematical Image Analysis Group, Department of Computer Science,
Saarland University, 66041 Saarbrücken

I hereby declare that this thesis is entirely the result of my own work except where otherwise indicated. I have used only the resources given in the list of references.

Yana Mileva,

Abstract

Variational methods currently belong to the most accurate techniques for the computation of the displacement field between the frames of an image sequence. Accuracy and time performance improvements of these methods are achieved every year. Most of the effort is directed towards finding better data and smoothness terms for the energy functional. Usually people are working mainly with black-and-white image sequences. In this thesis we will consider only colour images, as we believe that the colour itself carries much more information than the grey value and can help us for the better estimation of the optic flow. So far most of the research done in optic flow computation does not consider the presence of realistic illumination changes in the image sequences. One of the main goals of this thesis is to find new constancy assumptions for the data term, which overcome the problems of severe illumination changes. So far no research has been also done on combining variational methods with statistical moments for the purpose of optic flow computation. The second goal of this thesis is to investigate how and to what extend the optic flow methods can benefit from the rotational invariant moments. We will introduce a new variational methods framework that can combine all of the above mentioned new assumptions into a successful optic flow computation technique.

Acknowledgements

First of all, I would like to thank Prof. Joachim Weickert, who introduced me to the fascinating world of Image Analysis and Computer Vision and who, through his passion for the subject, convinced me go into this field.

I owe my highest gratitude to Dr. Andrés Bruhn for giving me the chance to work in the Mathematical Image Analysis Group at the Saarland University and for his help and supervision during the development of this thesis.

I also have to thank Dr. Martin Welk for his advice and the time he spent discussing some of the problems we had while working on the thesis.

I further want to thank the DAAD organization, for supporting my Master studies at the Saarland University, and especially Mrs. Gisela Yourid for all the help that she provided to me in times of bureaucratic difficulties.

I thank all of my friends for their support, and the wonderful time we had and will spend together.

And last but not least, I thank my parents for their ultimate support during all of these years.

Yana Mileva

Contents

Chapter 1

Introduction

1.1 Problem description

One of the key correspondence problems in Computer Vision is the computation of the optic flow field between two consecutive frames of an image sequence. An image sequence is an ordered set of images and we would like to be able to correctly estimate the direction and the velocity of the motion introduced in such an image sequence. In order to compute the motion we need to compare certain features that stay invariant in the sequence and that can help us to identify the objects and their movement in the scene. These features we call *invariance* or *constancy assumptions.*

So far, most of the research done in this field has been directed to computation of the flow field between the frames of black-and-white image sequences. In this thesis we investigate the computation of the optic flow between the frames of colour image sequences. Colour by itself carries more information about the objects than the grey value and thus, provides us more information that we can use for computing the flow.

There are two primary goals of this thesis: The first goal is to find new invariance assumptions that are valid for colour image sequences, where we have realistic illumination changes introduced. The second goal is to investigate whether and how the statistical moments can be used for optic flow computation. But let us first formulate the optic flow problem in a more formal manner.

1.2 What is Optic Flow?

The term "optic flow" refers to a visual phenomena that we experience every day. Essentially, optic flow is the apparent visual motion that we experience as we move through the world. Suppose you are sitting in a car or a train, and are looking out the window. You see trees, the ground, buildings, etc., appear to move backwards. This motion is optic flow.

Having a video (an image sequence) we would like to estimate the direction and the velocity of the motion in this video. For this purpose, we take two consecutive frames
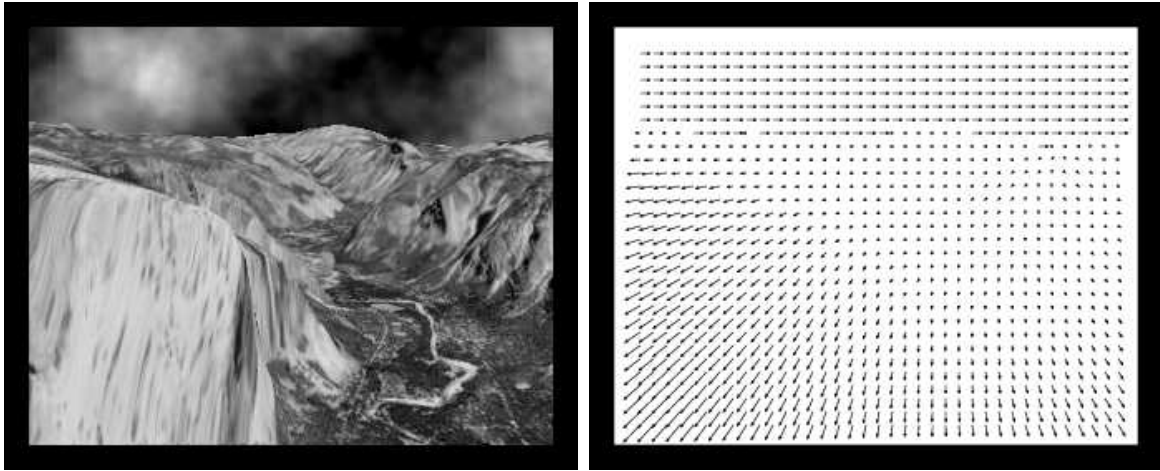
Figure 1.1: **Left:** The 9th frame of the famous Yosemite sequence; **Right:** The ground truth vector field, estimated between the 8th and the 9th frames of the same sequence;

of the sequence and try to estimate the position to which every pixel from the first frame has moved in the second frame.

The image is represented as a pixel matrix, where every pixel has a unique coordinate position. In order to represent the problem mathematically we consider a scalar-valued image sequence $f_0(x, y, t)$, where $(x, y)^T$ denotes the coordinates of the pixel at time t. We further consider that the temporal displacement between any two consecutive frames from the sequence is $\Delta t = 1$.

Let the image frame at time t be denoted by $f_0(x, y, t)$. Then the image frame at time $t+1$ will be denoted by $f_0(x + u, y + v, t + 1)$, where $x + u$ and $y + v$ show the new position of the pixel, which was previously at $(x, y)^T$ position. As we are working with real-world data, which is continuous, not discrete like the pixel notation, the actual displacement position coordinates are not necessary integer values.

Having formulated the problem like that, we can state that the computation of the optic flow is actually the computation of the dense vector field, where for every pixel we compute the vector $w = (u, v, 1)^T$ (see Figure 1.1).

1.3   Uses for Optic Flow

The optic flow has many fields of application.

One of the most popular applications of the optic flow is in robotics, where computing the optic flow for the image stream that comes through the robot's visual sensors helps the moving robot navigate in the space and avoid obstacles.

Many insects actually use optic flow to navigate in the air and this prompted the idea of using optic flow for navigation of unmanned aircrafts. Visual motion can be used to detect the relative change between speed and depth. This is useful for aircraft attempting to measure altitude.

Figure 1.2: Application fields of the optic flow: **Left:** Object tracking; **Center:** Human-machine interaction; **Right:** Driver Assistance system

A relatively similar way of using the optic flow is applied in the automobile industry (see Figure 1.2), where the car producers are integrating driver assistance systems which use optic flow to compute the motion around the car, thus preventing the car from having an accident by making it stop or just by warning the driver that a dangerous situation is approaching.

Human-machine interaction is another field where the optic flow can be used. The system is usually supplied with a video camera that records the human motion and then, based on this motion, makes its decisions. The computation of the human motion is done using optic flow algorithms.

Another well-known application is in video compression algorithms. Usually, in a real-world video sequence, not everything is moving, so it is not necessary to encode all of the information. Hence people usually split the sequence into frames and consider only a certain number of vital ones. For the rest of the frames, the optic flow is computed and stored, instead of whole frames.

These are only few examples that show how the optic flow can be used for solving different computer vision problems. Of course for every specific application there are specific problems that have to be addressed and taken into consideration. In the end the most important thing is to have a suitable optic flow computation algorithm that works for the particular application.

1.4 Outline of the Thesis

This thesis is organized as follows:

The variational methods are the most effective way, so far, to compute the optic flow. In Chapter 2 the general structure of the variational methods is briefly explained, together with the meanings of the data and the smoothness terms. We present two famous variational methods - the Horn-Schunck [17] and the Brox *et al.* methods [4].

As one of the main goals of this thesis is the computation of the optic flow for image sequences with introduced *realistic* illumination changes, in Chapter 3 we present a famous physical model for describing the light distribution. Later in the chapter,

we derive several photometric constancy assumptions, that remain valid even in the presence of severe illumination changes.

In Chapter 4 we discuss the statistical moments and how we can still retain rotational invariance, without having translational one. Eight rotational moments are derived and later used in the Experiments section.

In Chapter 5 new multichannel versions of the Horn-Schunck and the Brox *et al.* methods are developed. We need such extended frameworks, as we want to integrate several constancy assumptions at the same time in the data term and benefit from their positives.

We combine the theory from the above chapters in one and present and discuss our experimental results in Chapter 6.

This thesis is closed with conclusions and future work in Chapter 7.

Chapter 2

Variational Optic Flow Methods

Variational methods are among the best performing techniques for optic flow computation developed so far. These methods are global methods and, as such, they operate on the entire image domain. In essence, the variational optic flow methods recover the displacement field as the minimizer of a suitable energy functional. This energy functional consists of two terms - a *data term* and a *smoothness term*. The data term provides the information, while the smoothness term distributes this information. As we have these two terms, the computed flow field is always dense - in the cases where no useful information can be computed from the data term, the smoothness term fills in information from the neighbouring regions, so that in the end we have information for every single location of the image. More detailed explanations of the general structure of the variational methods will be provided in section 2.1.

The first who used variational methods for the purpose of optic flow computation were Horn and Schunck [17]. In section 2.2 we will present their method and will elaborate on the specific choice of the data and the smoothness term.

In the last years a lot of research has been done in order to construct refined data constraints and new and more efficient data terms [30, 3, 27, 28, 39, 4, 33, 5]. In section 2.3 we will present the method of Brox *et al.*, which is currently the best performing method available in the literature.

2.1 General Structure

If we are given the image sequence $f_0(x, y, t)$ and we would like to compute the flow field $w = (u, v, 1)^T$, variational methods tell us that we have to construct an energy functional that has the following structure:

$$E(w) = \int_{\Omega} \Big(D(f_0, w) + \alpha . S(\nabla f_0, \nabla w) \Big) \, dxdy \tag{2.1}$$

As you can see, the integration is performed over the entire image domain Ω, which characterizes the global methods (for more details regarding the local methods, please

refer to [24]). Depending on whether we work only with two frames from the sequence or we want to estimate the displacements in the whole sequence, this domain is spatial or spatio-temporal. In accordance, $\nabla = \nabla_2 = (\partial_x, \partial_y)^T$ if we work in the spatial domain and $\nabla = \nabla_3 = (\partial_x, \partial_y, \partial_t)^T$ if we work in the spatio-temporal domain. In this thesis we work only in the spatial domain.

Usually not the original image sequence f_0 is used, but a modified version f of it that has been convolved with a Gaussian kernel K_ρ with standard deviation ρ:

$$f = K_\rho * f_0 \tag{2.2}$$

This presmoothing step removes some noise, if such exists, and thus stabilizes the final result and makes f more suitable for operations like differentiation. The parameter ρ is an important parameter that can affect the final result of the optic flow computation.

Having the energy functional we are trying to minimize it and thus, penalizing deviations from the model assumptions, to find the best solution for our displacement field. The model assumptions are expressed by the *data term* $D(f, w)$ and the *smoothness* or *regularization term* $S(\nabla f, \nabla w)$. Each of these two terms has a specific role. The data term represents the assumption that certain features of the image do not change and remain constant over time, while the smoothness term reflects the claim that the neighbouring regions (pixels) belong to the same object and thus undergo a similar motion. The data term is responsible for generating the information for the optic flow and it stands for the constancy assumption(s) that are used. The smoothness term is responsible for the propagation of the computed information and the smoothing of the flow outliers. In case no information can be computed from the data term, the smoothness term has a filling-in effect, by filling in information at this spot, using the computed information from the neighbouring regions. This is where the parameter α will have an effect, controlling how much the smoothness term will prevail during the computation - the larger α is, the smoother the computed flow field is.

Minimization step: Now, having the energy functional, we have to minimize this functional, in order to find the best solution for the optic flow field.

If the constructed energy functional is strictly convex, it will have a unique solution $(u, v)^T$ that minimizes it. This solution is also the solution of the Euler-Lagrange equations, as proved in [12].

If we have the energy function

$$E(w) = \int_\Omega F(f, w, \nabla f, \nabla w) \, dxdy \tag{2.3}$$

where $F(f, w, \nabla f, \nabla w) = D(f, w) + \alpha.S(\nabla f, \nabla w)$ from (2.1), then the corresponding Euler-Lagrange equations look like this:

$$\partial_x F_{u_x} + \partial_y F_{u_y} - F_u = 0 \tag{2.4}$$

$$\partial_x F_{v_x} + \partial_y F_{v_y} - F_v = 0$$

So in order for us to minimize our energy functional (2.1) we have to solve its Euler-Lagrange equations. And according to (2.4) these equations look like this:

$$\alpha . \partial_x S_{u_x} + \alpha . \partial_y S_{u_y} - D_u = 0 \tag{2.5}$$

$$\alpha . \partial_x S_{v_2} + \alpha . \partial_y S_{v_y} - D_v = 0$$

with homogeneous Neumann boundary conditions:

$$n^T \left(\begin{array}{c} \alpha . S_{u_x} \\ \alpha . S_{u_y} \end{array} \right) = 0, \ n^T \left(\begin{array}{c} \alpha . S_{v_x} \\ \alpha . S_{v_y} \end{array} \right) = 0 \tag{2.6}$$

where n is the unit normal vector on $\partial\Omega$.

Discretization step: As we are working with discrete images, where the integration domain Ω is discretized by a grid of $N_1 \times N_2$ pixels, the flow field that we are looking for will be the solution of the above system of two partial differential equations (2.5) in their discretized variants. To discretize these PDEs we have to approximate the derivatives by finite differences. The result of this discretization is a linear or non-linear system of equations that has to be solved by a suitable numerical algorithm. In our implementation we are using the Successive Over Relaxation (SOR) method [15]. For a further discussion regarding different types of system solvers, please refer to [6].

Variational formulations of the optic flow estimation problem have several advantages that can be formulated as follows [6]:

- **Transparency:** These formulations lead to a sound model where one clearly has to formulate the model assumptions.

- **Numerics:** These methods make use of the long experience of using numerical mathematics to solve partly difficult optimization problems.

- **Motion types:** The models are suitable for all types of motion - not only translatory one.

- **Flexibility:** The constancy assumptions can be easily modified according to the application.

- **Filling-in effect:** When there is no reliable estimation possible we can profit from the filling-in effect at these locations.

2.2   The Horn-Schunck method

Horn and Schunck were the first people who decided to use variational methods for the computation of the optic flow field. The Horn-Schunck method is a global method that introduces a global constraint of smoothness over the computed flow field.

Let us now take a closer look at this functional and the way it has been constructed.

The energy functional that they constructed looks like this:

$$E(w) = \int_\Omega \left((f_x u + f_y v + f_t)^2 + \alpha(|\nabla u|^2 + |\nabla v|^2) \right) dxdy \qquad (2.7)$$

where f_x denotes the partial derivative of f with respect to x and $|\nabla u| = \sqrt{u_x^2 + u_y^2}$

Let us now take a closer look at this functional and the way it has been constructed.

Data term   The data term in this functional is in fact the famous grey value constancy assumption. Horn and Schunck stated that, if the displacement between the frames is small enough, the grey value will not drastically change from frame to frame. So, for their approach, they considered that the grey value actually remains constant over time. This statement can be mathematically expressed as follows:

$$f(x, y, t) = f(x + u, y + v, t + 1) \qquad (2.8)$$

or

$$f(x + u, y + v, t + 1) - f(x, y, t) = 0 \qquad (2.9)$$

This literally means that if $f(x, y, t)$ is the grey value of pixel (x, y) at time t, this value remains equal to the grey value of pixel $(x + u, y + v)$ at time $t + 1$. Note that if we consider that $f(x, y, t)$ is another feature of the image that remains constant over time, this statement is still valid. This will be illustrated in the next section.

If we consider that the displacements between the frames are small we can linearize this constancy assumption using first order Taylor expansion:

$$f(x + u, y + v, t) = f(x, y, t) - f_x u - f_y v - f_t \qquad (2.10)$$

And from equation (2.9) we can say that:

$$f_x u + f_y v + f_t = 0 \qquad (2.11)$$

which is exactly what Horn and Schunck used for their data term, as can be seen from (2.7). The assumption is squared, because they want to penalize all possible deviations from zero that may appear.

Please note that, this linearization is valid only under the condition that we have small displacements. In the case of large displacements it will be violated and thus a more sophisticated approach, like the Brox *et al.* method, will be needed.

Smoothness term Unfortunately, it is not possible to uniquely determine u and v only from the above stated equation (2.11). This problem is known in the literature as the aperture problem [2]. The method Horn and Schunck used to combat this problem is by adding an additional assumption, namely, they consider that the computed optic flow field is globally smooth. This they have expressed by the smoothness term in (2.7) (the so called *homogeneous regularizer*): $|\nabla u|^2 + |\nabla v|^2$.

This type of smoothness term though is not always the most suitable one, as it considers that the resulting flow field has no discontinuities, which in most of the cases is not true (the objects' boundaries should be preserved and not smoothed). In the following section a more suitable smoothness term will be presented.

The Horn-Schunck energy functional actually introduces two assumptions:

1. The grey value constancy assumption, which imposes constancy over the grey value image feature and tells us that the grey value doesn't change over time.

2. The flow field smoothness assumption, which imposes additional constraints on the solution and tells us that the flow field that we are trying to compute has to be globally smooth.

Now we can easily construct the Euler-Lagrange equations for this energy functional based on (2.4):

$$\alpha u_{xx} + \alpha u_{yy} - f_x(f_x u + f_y v + f_t) = 0 \tag{2.12}$$

$$\alpha v_{xx} + \alpha v_{yy} - f_y(f_x u + f_y v + f_t) = 0 \tag{2.13}$$

or, as usually expressed:

$$f_x(f_x u + f_y v + f_t) - \alpha \Delta u = 0 \tag{2.14}$$

$$f_y(f_x u + f_y v + f_t) - \alpha \Delta v = 0 \tag{2.15}$$

where $\Delta u = u_{xx} + u_{yy}$ and $\Delta v = v_{xx} + v_{yy}$.

After discretizing these equations, what remains is to solve a system of equations, which is obviously a linear one, and for which there are plenty of solvers available.

2.3 The Brox *et al.* method

In this section, we discuss the famous Brox *et al.* method, which is described in [4]. This is much more sophisticated method than the Horn-Schunck method and accordingly it gives much better results.

The Brox *et al.* method is a novel approach that integrates several successful concepts and addresses all the problems of the Horn-Schunck model.

The energy functional proposed in [4] for the 2-D case looks like this:

$$E(w) = \int_\Omega \Big(D(f, w) + \alpha S(\nabla f, \nabla w)\Big)\, dxdy \qquad (2.16)$$

where

$$D(w) = \Psi\Big(D_{brightness}(f, w) + \gamma D_{gradient}(f, w)\Big) \qquad (2.17)$$

and

$$S(\nabla f, \nabla w) = \Psi\Big(|\nabla u|^2 + |\nabla v|^2\Big) \qquad (2.18)$$

with

$$D_{brightness}(f, w) = |f(x + u, y + v, t + 1) - f(x, y, t)|^2 \qquad (2.19)$$

$$D_{gradient}(f, w) = |\nabla f(x + u, y + v, t + 1) - \nabla f(x, y, t)|^2 \qquad (2.20)$$

Now let us take a closer look at this energy functional and try to explain the meaning of the data term and the smoothness term.

Data term The data term is a sum of two constancy assumptions - the famous brightness constancy assumption and the gradient constancy assumption. An additional parameter γ is added to weight the importance of the assumptions. However, as you can see from (2.19) and (2.20), these two assumptions are not linearized, but instead the authors use the original assumptions, where they consider that the grey value remains unchanged

$$f(x + u, y + v, t + 1) = f(x, y, t) \qquad (2.21)$$

and the value of the gradient remains unchanged

$$\nabla f(x + u, y + v, t + 1) = \nabla f(x, y, t) \qquad (2.22)$$

The linearization step that Horn and Schunck took is only valid under the assumption that the image changes linearly along the displacement, which in general is not the case, especially where there are large displacements. Hence the authors restrain from linearizing the assumptions.

The authors also decided to use not only the grey-value assumption, but also the gradient constancy assumption. The grey-value assumption is easy to violate when even slight changes in brightness occur. In comparison, the gradient assumption is invariant under grey value changes. From another point of view, the gradient constancy assumption is useful when translatory motion is introduced, while the grey-value assumption is better for more complicated motion patterns. The combination of these two assumptions proved to give impressive results [4].

Another further improvement that has been introduced is the use of the Ψ function. Horn and Schunck are minimizing the squared constancy assumption, but using such

a quadratic penalizer gives too much influence to the outliers that may appear in the image. So, in this respect, it is better to use another kind of penalizing function and the function that Brox *et al.* are using is $\Psi(s) = \sqrt{s^2 + \varepsilon^2}$. The parameter ε can be easily set to a fixed value, 0.001 for example. Thus $\Psi(s)$ is still a convex function, which is easier to minimize, and at the same time does not introduce additional parameters and penalizes the outliers in a less severe way.

Smoothness term   If we look at the smoothness term that Horn and Schunck use, we will notice that they are imposing the same smoothness constraint in all directions, meaning that they are smoothing in the same manner everywhere and doing this they actually smooth all discontinuities that may appear in the flow field. In reality though, we do want to see discontinuities in the computed optic flow, when these discontinuities correspond to the objects in the image. That is why the authors are imposing *piecewise* smoothness, penalizing the variation from the flow field, using the same Ψ function from above and thus receiving more accurate and realistic results.

Now, having the energy functional (2.16), we have to minimize it. For this purpose we have to construct the Euler-Lagrange equations (2.4). Here the computation is a bit more complicated, so we will make it step by step.

We have the following function (see (2.16)):

$$F(f, w, \nabla f, \nabla w) = D(w) + \alpha S(w)$$

where we have that

$$F = \Psi\Big(|f(x+u, y+v, t+1) - f(x, y, t)|^2 + \gamma|\nabla f(x+u, y+v, t+1) - \nabla f(x, y, t)|^2\Big) + \alpha\Psi\Big(|\nabla u|^2 + |\nabla v|^2\Big)$$

and we have to compute $\partial_x F_{u_x}, \partial_y F_{u_y}, \partial_x F_{v_x}, \partial_y F_{v_y}, F_u$ and F_v in order to have our Euler-Lagrange equations.

$$\partial_x F_{u_x} = 2\alpha\, \partial_x \Psi'(|\nabla u|^2 + |\nabla v|^2)u_x + 2\alpha\, \Psi'(|\nabla u|^2 + |\nabla v|^2)u_{xx}$$

$$\partial_y F_{u_y} = 2\alpha\, \partial_y \Psi'(|\nabla u|^2 + |\nabla v|^2)u_y + 2\alpha\, \Psi'(|\nabla u|^2 + |\nabla v|^2)u_{yy}$$

$$\partial_x F_{v_x} = 2\alpha\, \partial_x \Psi'(|\nabla u|^2 + |\nabla v|^2)v_x + 2\alpha\, \Psi'(|\nabla u|^2 + |\nabla v|^2)v_{xx}$$

$$\partial_y F_{v_y} = 2\alpha\, \partial_y \Psi'(|\nabla u|^2 + |\nabla v|^2)v_y + 2\alpha\, \Psi'(|\nabla u|^2 + |\nabla v|^2)v_{yy}$$

$$
\begin{aligned}
F_u \;=\; & \Psi'\Big(|f(x+u,y+v,t+1)-f(x,y,t)|^2 + \gamma|\nabla f(x+u,y+v,t+1)-\nabla f(x,y,t)|^2\Big). \\
& \Big[2\,|f(x+u,y+v,t+1)-f(x,y,t)|\,\partial_u f(x+u,y+v,t+1) - \\
& \gamma\,2\,|\nabla f(x+u,y+v,t+1)-\nabla f(x,y,t)|\,\partial_u \nabla f(x+u,y+v,t+1)\Big] \\
F_v \;=\; & \Psi'\Big(|f(x+u,y+v,t+1)-f(x,y,t)|^2 + \gamma|\nabla f(x+u,y+v,t+1)-\nabla f(x,y,t)|^2\Big). \\
& \Big[2\,|f(x+u,y+v,t+1)-f(x,y,t)|\,\partial_v f(x+u,y+v,t+1) - \\
& \gamma\,2\,|\nabla f(x+u,y+v,t+1)-\nabla f(x,y,t)|\,\partial_v \nabla f(x+u,y+v,t+1)\Big]
\end{aligned}
$$

Now, having made the calculations above, we can construct the Euler-Lagrange equations and, after some small modifications, they will look like this:

$$
\begin{aligned}
0 \;&=\; \partial_x F_{u_x} + \partial_y F_{u_y} - F_u & (2.23) \\
&=\; F_u - \alpha\,div\Big(\Psi'(|\nabla u|^2 + |\nabla v|^2).\nabla u\Big) \\
0 \;&=\; \partial_x F_{v_x} + \partial_y F_{v_y} - F_v & (2.24) \\
&=\; F_v - \alpha\,div\Big(\Psi'(|\nabla u|^2 + |\nabla v|^2).\nabla v\Big)
\end{aligned}
$$

After discretizing the above equations, what remains is to solve the resulting system of equations. Unfortunately, it is not as easy in this case as in the Horn-Schunck method, because here we have a non-linear system of equations. Moreover as the variational model leads to a non-convex optimization problem, a well-elaborated minimization scheme becomes necessary. In order to find a good minimizer of the original non-linear model, all linearisations are consequently postponed to the numerical scheme. While linearisations in the model would immediately compromise the overall performance of the system, linearisations in the numerical scheme can help to improve convergence to the global minimum. This is the basic idea that stands behind the minimization strategy of the Brox *et al.* method. In our experiments, based on this method, we are again using the SOR approach, which can be adapted to solving non-linear systems of equations. In general the system of equations that has to be solved is fairly big and sparse and takes long time to be solved. If we require real-time performance, a faster and a more sophisticated method, like multigrid, should be used (see [6, 9, 7, 8] for further details).

In case we have displacements that are larger than one pixel per frame, there is a danger that our method will be trapped at some local minimum and will not find the global minimum and thus the desired solution. The Brox *et al.* method offers a solution to this problem, too. The authors are proposing the use of multiple scale levels with the coarse-to-fine strategy [6, 4, 33]. So, in the case where we have large displacements, we start solving a coarser version of the problem. This newly defined problem will hopefully have a unique solution and this solution will be close to the

global minimum of the original problem. The coarse solution is used as initialization for solving a refined version of the problem until we finally reach the original scale of the problem and have our global minimum solution.

2.4 Summary

In this chapter we have presented the general structure of the variational methods for optic flow computation and have briefly explained the way the energy functional is constructed, as well as the meaning of the data and the smoothness terms.

The general structure of these methods have been illustrated later in the chapter with two famous variational methods. The method of Horn-Schunck is the first and the basic variational method and as such we have used it to illustrate the structure of the energy functional. Later we have briefly presented the much more sophisticated method of Brox *et al.*, which has the same structure for the energy functional, but manages to overcome all of the problems that may appear from the Horn-Shunck method. The Brox *et al.* method is based on the coarse-to-fine strategy, which helps to avoid local minima; on non-linearized constancy assumptions, which makes is suitable for computing the flow even in the presence of large displacements and also on the piecewise type of smoothness imposed on the flow field solution.

Later in Chapters 5 and 6 we will again follow the same strategy for theoretical explanations, deriving them first from the simple Horn-Schunck method, and later extending them to the sophisticated Brox *et al.* method.

Chapter 3

Illumination Invariants

One of the main goals of this thesis is to deal with realistic illumination changes. The presence of illumination changes can be crucial for the computation of the optic flow. Let us assume for a moment that the second frame is brighter than the first one, due to some atmospheric conditions or automatic adjustments of the camera. In this case the brightness constancy assumption, for example, will fail completely, as every pixel from the second frame will actually be brighter. What we want to do is to find constancy assumptions that will remain valid even under severe illumination changes.

We are mainly interested in realistic illumination changes for colour image sequences. So far, most of the research done in optic flow computation, has been based on black-and-white images. In reality though, we usually have colour images. The colour itself carries more information than the grey value and we can extract more useful data from it.

In this chapter we will present the dichromatic reflection model, which mathematically describes the physical colour of the objects, taking into account the illumination light, the object surface and the body reflectance. Based on this model three types of photometric invariants can be described, as the reader will see later in the chapter.

The main contribution of this thesis is explained in section 3.2, where we will present different photometric constancy assumptions, suitable for different illumination conditions.

3.1   The Dichromatic Reflection Model

One of the most important characteristics of an object is its colour. Unfortunately, it is not easy to model and represent colour mathematically, because there are too many things that can influence the colour that the viewer sees. Nevertheless, people are trying to find a suitable model and one of the most widely used models so far is the dichromatic reflection model [21, 23, 36, 22, 13].

Let us consider a surface that is illuminated by light. This light is characterized by some power distribution e. The reflected light has its own power distribution, which is not necessary equal to that of the incident light. The dichromatic reflection model is
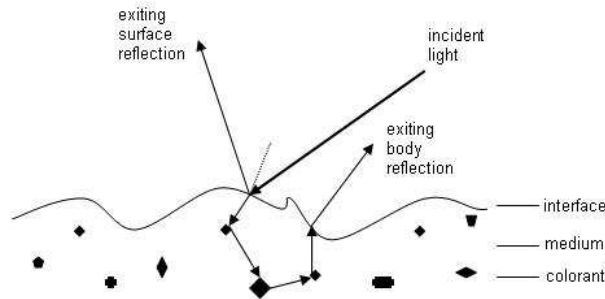
Figure 3.1: Light reflection of dielectric materials

especially developed to describe the observed colour $L(X)$ at a certain surface location $X = (x,y)^T$ of an inhomogeneous dielectric material [21].

A dielectric material is a material that cannot conduct electricity. Most of these materials are inhomogeneous. They consist of a medium and some embedded pigment particles, as it is shown in Figure 3.1. The medium is generally transparent and it transmits the light beams without changing their colour or direction. The pigment particles though can be hit by a light beam and thus they can absorb or scatter it.

When a light beam hits a dielectric material, the reflected light that is visible for an observer is composed of two parts: the light reflected from the material's surface (highlights and gloss on the object) and the light reflected by the material's body (object colour and object shading).

3.1.1   Reflection of the surface

If the surface of the body is completely smooth, we will have a mirror-like reflection, which means that the whole amount of the incidence light will be reflected. If you look at Figure 3.2, you can see that in this case the incident light beam, the exiting light beam and the surface normal will lie on the same plane. However, the surfaces of many materials are not that smooth and have some rough texture. In these cases the direction of the reflected light depends on the orientation of the local surface normals. Then the incidence light is scattered to some degree around the global angle of the perfect mirror reflection.

Here we note that matte materials don't have surface reflection component.

3.1.2   Reflection of the body

For dielectric materials, not all of the incidence light is reflected by the surface, so some percentage of it penetrates the body of the object. The light that goes into the body travels unaltered until it hits a pigment particle. After it hits a particle, part of the light gets absorbed by this particle, part of it passes through the particle and part of it gets scattered. At some point the remaining light returns to the surface and is
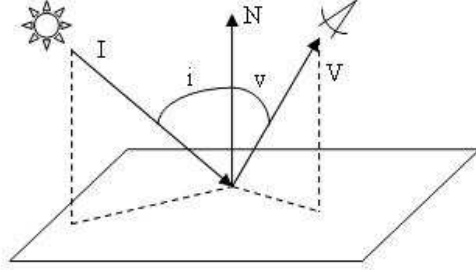
Figure 3.2: Photometric angles

emitted.

This reflection process is called body reflection. The body reflection provides the characteristic object colour. Its geometric and photometric properties depend on many factors, such as: the transmitting properties of the body medium; the scattering and absorption properties of the pigment particles; and the shape and the distribution of these particles. The opacity of the material also depends on the pigmenting particles, specifically on their distribution and size (a material is opaque if no part of the refracted light beam is able to penetrate through the entire material).

3.1.3 The Dichromatic Reflection Model

So far we have shown how complex the path of the light is. Now we will present the dichromatic reflection model [21, 23, 36, 22, 34], which is a mathematical model that approximates the reflection process of dielectric, inhomogeneous, opaque materials.

As this model is just an approximation of the real light function, we have to make a number of assumptions. First of all we assume uniform colour of illumination from a single light source. Then we assume that the pigment particles are randomly distributed and that they are entirely into the body (and not on its surface). This allows us to state that we have a single spectrum of body reflection and a single spectrum of surface reflection. Despite these assumptions, this model is still a very good approximation of the real light reflection, as proved in [21].

According to the dichromatic reflection model, the light L that is reflected from an object point can be represented as a sum of the light reflected from the surface L_s and the light reflected from the body L_b:

$$L(\lambda, i, v, g) = L_s(\lambda, i, v, g) + L_b(\lambda, i, v, g) \tag{3.1}$$

The reflected light L is a function of the wavelengths of the light spectrum λ and the angles from Figure 3.2, i, v and g, where i is the angle between the incidence light beam I and the surface normal at this point N (the incidence angle); v is the angle between the normal N and the viewing direction V (the viewing angle) and g is the angle between I and V (the phase angle).

Based on our assumptions we can state that we have a single spectrum of body reflection and a single spectrum of surface reflection, which allows us to separate the spectral reflection properties $c(\lambda)$ from the geometric reflection properties $m(i, v, g)$. The properties $c(\lambda)$ remain unaltered as the geometric angles i, v, g change. Thus we can state that:

$$L_s(\lambda, i, v, g) = c_s(\lambda)\, m_s(i, v, g) \tag{3.2}$$

$$L_b(\lambda, i, v, g) = c_b(\lambda)\, m_b(i, v, g) \tag{3.3}$$

Let us consider $e(\lambda)$ to be the spectral power of the incident light. As we have assumed uniform colour of illumination, we can state that the spectral distribution e is independent of the image coordinate. Thus we can state that the dichromatic reflection model is described by the following equation [21]:

$$L(\lambda, i, v, g) = e\Big(c_s(\lambda)\, m_s(i, v, g) + c_b(\lambda)\, m_b(i, v, g)\Big) \tag{3.4}$$

We thus describe the light that is reflected from an object's point, and hence the colour that the viewer sees, as a mixture of two distinct spectral power distributions, c_s and c_b, each of which is scaled according to the geometric reflection properties of the surface and body reflections m_s and m_b.

We also note that the dichromatic reflection model equation simplifies to [38]:

$$L(\lambda, i, v, g) = e\, c_b(\lambda)\, m_b(i, v, g) \tag{3.5}$$

for objects with matte surfaces, i.e. for objects that have no surface reflection.

So far we have defined the dichromatic reflection model that describes the physical light reflection process. This model is defined for all wavelengths λ. However, when working with real colour images, we do not work with wavelengths, but with the R, G and B channels. Thus we reduce the incoming light beam to a RGB colour triple. We perform a spectral integration over the entire visible spectrum and thus transform [20, 35] equation (3.4) into:

$$C(X) = e\Big(C_s(X)\, m_s + C_b(X)\, m_b\Big) \tag{3.6}$$

where $C(X) = (R(X), G(X), B(X))^T$ is the observed RGB colour at image location $X = (x, y)^T$ and C_s and C_b are also a RGB triple and represent the interface and the body reflection colour.

The way we have modified equation (3.4) we can also modify equation (3.5):

$$C(X) = e\, C_b(X)\, m_b \tag{3.7}$$

As we are working with colour images, we will consider equations (3.6) and (3.7) in the following parts of this chapter.

Figure 3.3: **Left:** Frame 10 of the "Street" sequence; **Right:** Frame 11

3.2 Photometric Invariants

Now let us look at how we can model our constancy assumptions, using this dichromatic reflection model.

3.2.1 Types of Photometric Invariants

We already know how to represent the colour at a single location of an object. With respect to the dichromatic reflection model we have three photometric variables that can change, namely e, m_b and m_s. So, we can state that we have the following three types of photometric invariants:

- **Type 1:** Invariants with respect to the light source intensity e. These are invariants under *global multiplicative* illumination changes.

- **Type 2:** Invariants with respect to both e and m_b. These are invariants with respect to *local multiplicative* illumination changes, which result in *shadow* and *shading* changes in the image.

- **Type 3:** Invariants with respect to e, m_b and m_s. These are invariants with respect to both *local multiplicative* and *local additive* illumination changes and result in *highlights* and *specular reflections* in the image.

So far we have classified, according to the dichromatic reflection model, the photometric invariants that may possibly exist in our image. Let us now take a look at some concrete invariant assumptions.

The constancy assumptions that are to be presented are organized in three groups based on the applied strategy: normalization of the channels; taking the spacial derivatives of the channels; and considering channels from different colour spaces.

Figure 3.4: **Left:** the red channel for frame 10 of the "Street" sequence; **Center:** the green channel; **Right:** the blue channel

3.2.2   The RGB constancy assumption

When we work with black-and-white images, the most widely used constancy assumption is the grey-value constancy. With the grey-value constancy we assume that the grey value does not change over time and we express this with the equation:

$$f(x, y, t) = f(x + u, y + v, t + 1) \tag{3.8}$$

When it comes to colour images, we can re-formulate this assumption and in this case it will actually state that the *colour* remains the same over time. As the colour is a combination of the red(R), green(G) and blue(B) channels, our assumptions will look like this:

$$R(x, y, t) = R(x + u, y + v, t + 1) \tag{3.9}$$

$$G(x, y, t) = G(x + u, y + v, t + 1)$$

$$B(x, y, t) = B(x + u, y + v, t + 1)$$

Usually the RGB channels, when extracted from an image, are in the range $[0, 255]$. The range can be easily normalized to $[0, 1]$, but here we prefer to use the original RGB values.

Figure 3.4 shows exactly how these three channels look, based on frame 10 of the "Street" colour image sequence (Figure 3.3).

Of course, as in the case of the grey-value constancy assumption, the RGB channels assumptions are not valid under changing illumination conditions. So, we have to do something in order to make them invariant.

3.2.3   Normalized RGB assumptions

One thing that we can do in order to make the available RGB channels invariant is to normalize them. So, in this case our assumptions will be formulated as:

$$\frac{R(x,y,t)}{N(x,y,t)} = \frac{R(x+u,y+v,t+1)}{N(x+u,y+v,t+1)}, \tag{3.10}$$

$$\frac{G(x,y,t)}{N(x,y,t)} = \frac{G(x+u,y+v,t+1)}{N(x+u,y+v,t+1)},$$

$$\frac{B(x,y,t)}{N(x,y,t)} = \frac{B(x+u,y+v,t+1)}{N(x+u,y+v,t+1)}$$

We propose two types of normalization that are often used in the literature [13, 10, 11], normalization by:

- arithmetic mean, with $N = R + G + B$, and

- geometric mean, with $N = \sqrt[3]{RGB}$

As we are performing normalization, the normalized RGB channels are, of course, in the $[0,1]$ range. If we substitute the dichromatic reflection model equation (3.7) into the above equations, we see that these assumptions are actually invariant under local multiplicative illumination changes. This classifies these assumptions as Type 2 assumptions.

3.2.4   Derivatives assumptions

Another strategy that can be applied is the computation of spatial derivatives [37, 4, 33]. The assumptions that can be made with respect to these derivatives state that the value of the derivatives remains the same. This can be mathematically expressed as follows:

$$R_x(x,y,t) = R_x(x+u,y+v,t+1) \tag{3.11}$$

$$G_x(x,y,t) = G_x(x+u,y+v,t+1)$$

$$B_x(x,y,t) = B_x(x+u,y+v,t+1)$$

$$R_y(x,y,t) = R_y(x+u,y+v,t+1)$$

$$G_y(x,y,t) = G_y(x+u,y+v,t+1)$$

$$B_y(x,y,t) = B_y(x+u,y+v,t+1)$$

where R_x, G_x, B_x are the spatial derivatives of the R, G and B channels with respect to x and R_y, G_y, B_y are the spatial derivatives with respect to y.

Figure 3.5 demonstrates the effects of these derivative assumptions.

If we substitute the dichromatic reflection model equation (3.6) into the above assumptions, we see that these assumptions are not invariant under any illumination
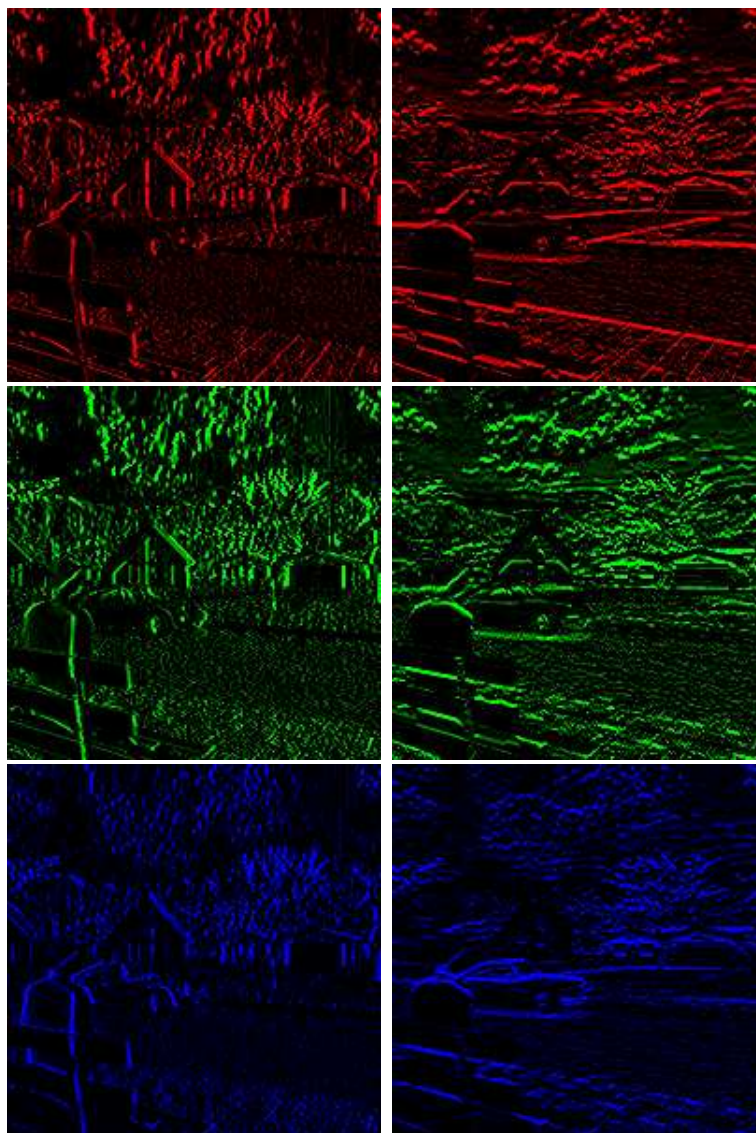
Figure 3.5: **Top Left:** the R_x channel for frame 10 of the "Street" sequence; **Top Right:** the R_y channel; **Middle Right:** the G_x channel; **Middle Left:** the G_y channel; **Bottom Left:** the B_x channel; **Bottom Right:** the B_y channel;
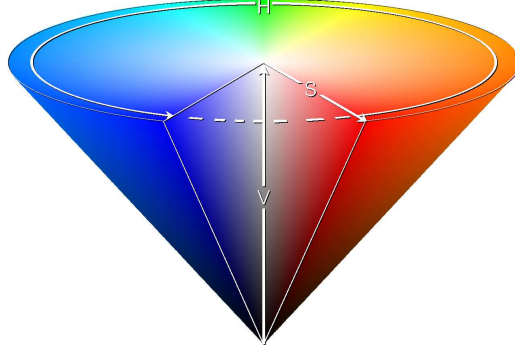
Figure 3.6: Visual representation of the HSV colour space [41]

changes. As such they are not useful to us. But, by applying some 'tricks', we can change this.

We can make them invariant by, instead of taking the derivatives of the R, G and B channels, we take the derivatives of the $ln(R), ln(G)$ and $ln(B)$ channels. Using this strategy we have the following **log-derivatives** constancy assumptions:

$$(lnR)_x(x, y, t) = (lnR)_x(x + u, y + v, t + 1) \tag{3.12}$$

$$(lnG)_x(x, y, t) = (lnG)_x(x + u, y + v, t + 1)$$

$$(lnB)_x(x, y, t) = (lnB)_x(x + u, y + v, t + 1)$$

$$(lnR)_y(x, y, t) = (lnR)_y(x + u, y + v, t + 1)$$

$$(lnG)_y(x, y, t) = (lnG)_y(x + u, y + v, t + 1)$$

$$(lnB)_y(x, y, t) = (lnB)_y(x + u, y + v, t + 1)$$

In this way we have gained invariance with respect to the light source intensity e, i.e with respect to global multiplicative changes, as can be seen by substituting the dichromatic reflection model equation (3.6) into the above assumptions. This makes these assumptions Type 1 assumptions.

3.2.5 Colour Spaces

So far we have worked only with the R, G and B channels and have tried to modify them in such a way as to make them invariant under some type of illumination changes. Now we will turn our attention towards other colour spaces, which can offer us different properties from those in the RGB space. We will take a look at some transformations of the RGB space.

Figure 3.7: The HSV channels - black-and-white coding. **Left:** the Hue channel for frame 10 of the "Street" sequence; **Center:** the Saturation channel; Bottom: the Brightness channel

The HSV colour space

The HSV colour space is a conical transformation of the RGB space, where H stands for *hue*, S stands for *saturation* and V stands for *value/intensity*. Figure 3.6 presents the HSV space and the hue, saturation and value axes.

The RGB colour space is based on the RGB colour model, where the R, G and B channels can be combined in various ways in order to produce the final colour. In contrast, the three channels of the HSV space depict colour in a different manner. HSV encapsulates information about a colour in terms that are more familiar to humans. The hue is that aspect of a colour described with names such as "red", "blue", "yellow", etc. In painting colour theory, the hue refers to the pure colour, one without added white or black pigment. Saturation (or purity) refers to the intensity of a specific hue. A highly saturated hue has a vivid, intense colour, while a less saturated hue appears more muted and grey. With no saturation at all, the hue becomes a shade of grey. The brightness, accordingly, describes how close a colour is to white or black, i.e. it depicts how bright a colour is. We have to note that the $RGB \rightarrow HSV$ mapping is not a one-to-one mapping, as there is not a unique HSV correspondence for every RGB colour. This could cause troubles during the computation. What we have decided to do is just to give value 0 to the hue, when the hue is not defined.

Now let us take a look at the $RGB \rightarrow HSV$ transformation [38]:

$$h = arctan\Big(\frac{\sqrt{3}(R-G)}{R+G-2B}\Big)$$

$$s = \sqrt{\frac{3(R-G)^2 + (R+G-2B)^2}{6}}$$
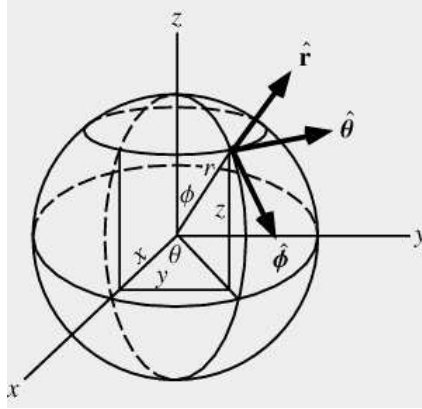
$$v = \frac{R+G+B}{\sqrt{3}}$$

Figure 3.8: Spherical Transformation [25]

If we have the RGB values in the range $[0, 1]$, we will have the following ranges for the hue, saturation and value: $h \in [0^0, 360^0], s \in [0, 1], v \in [0, 1]$. If we have the RGB in the $[0, 255]$ range, we have $h \in [0^0, 360^0], s \in [0, 255]$ and $v \in [0, 255]$. Please note that the hue is measured in degrees.

Figure 3.7 shows each of these three channels.

The hue is a well known photometric invariant with respect to both shading and highlights, which makes it a Type 3 invariant; this is easily seen if we substitute equation (3.6) into the hue equation. The reader should note that the hue channel, in comparison with the rest of the presented channels, is the one that is most invariant against illumination changes, but at the same time this invariance comes at a cost. By the computation of the hue channel, we are throwing away a lot of information, as we are preserving only the information that is invariant against both shading and highlights. So, even if the hue looks like a universal invariant, it should be used carefully. It may happen that in some cases it is better to combine two less invariant assumptions rather than use the hue invariance.

In contrast to the hue, the saturation is invariant only with respect to shadow and shading changes, which makes it a Type 2 invariant.

The brightness channel is not invariant at all. Thus it shall not be used as a constancy assumption.

The $\tau\phi\theta$ colour space

In the previous section we have presented a conical transformation of the RGB space into the HSV space. Here we will present a spherical transformation, namely the transformation from the RGB into the $\tau\phi\theta$ space. An illustration of the spherical transformation can be found on Figure 3.8. The $RGB \rightarrow \tau\phi\theta$ transformation is given by the following equations [38]:

Figure 3.9: The $\tau\phi\theta$ channels: black-and-white coding. **Left:** the τ channel for frame 10 of the "Street" sequence; **Center:** the ϕ channel; **Right:** the θ channel.

$$\tau = \sqrt{R^2 + G^2 + B^2}$$

$$\phi = arctan\left(\frac{G}{B}\right)$$

$$\theta = arcsin\left(\frac{\sqrt{R^2 + G^2}}{\sqrt{R^2 + G^2 + B^2}}\right)$$

Here, τ denotes the magnitude of the colour vector and ϕ and θ are the longitude and latitude angles. With respect to the above definition, these channels vary in the following range: $\tau > 0; \phi \in [0^0, 360^0]; \theta \in [0^0, 180^0]$.

In Figure 3.9 we have demonstrated how these channels look.

As one can see, after substituting in equation (3.7), the ϕ and θ channels are invariant with respect to shadow and shading and this makes them Type 2 invariants.

The τ channel is not invariant.

3.2.6 Combining Different Assumptions

The assumptions presented so far are valid for different illumination condition changes. Of course, in one image sequence we may have not only one type, but several types of illumination changes. In this case we can combine some constancy assumptions that are valid under some conditions with assumptions that are valid under other conditions. In this way we can gain invariance against different illumination changes.

After we have decided to combine several assumptions we could also decide to give more importance to one of the assumptions than to the others. This we can do by assigning a weight variable to every assumption and giving values to these variables, in accordance with our intentions. This will be illustrated in the multichannel models from Chapter 5.

3.3   Summary

In this chapter we have presented several constancy assumptions that are invariant against three types of illumination changes. These three types of illumination changes we have derived from the dichromatic reflection model equation.

The presented constancy assumptions are mainly based on one of the following three strategies: normalization of the channels, taking the spatial derivatives of the channels and using channels from different colour spaces. Each one of these assumptions is constant against some type of illumination changes.

In order to use these assumptions for optic flow computation we need a variational method that integrates multiple channels into the data term. In Chapter 5 we modify the Horn-Schunck and the Brox *et al.* methods so that they can be used for multiple constancy assumptions.

Chapter 4

Moment Invariants

In this chapter we will present a novel approach for determining constancy assumptions based on the statistical moments. Until now, the constancy assumptions have been based on the principle that certain properties of the pixels remain unchanged between two consecutive frames of an image sequence. In the previous chapters we have presented some of these properties, such as the grey value for black-and-white images and the hue channel for colour images. In this chapter we will present the moments constancy assumptions, which are based on different principles. In the literature, the moments are most frequently used for object and pattern recognition. Comparison of two objects means determining whether they belong to the same class of object or not. As some objects may undergo several transformations such as translations, scalings and rotations, the used moments have to be invariant under such transformations.

In the first part of this chapter we will present the moments and the modifications that are made to them in order to be invariant under the above mentioned transformations.

In the second part of the chapter we will present the moments constancy assumptions that we have developed for optic flow computation.

4.1   Moments

The concept of moment in mathematics evolved from the concept of moment in physics. Let us suppose that we are given a discrete image f, where (i, j) denote the coordinates of the $(i, j)^{th}$ pixel. Then, the value of the $(i, j)^{th}$ pixel is given by $f_{i,j}$.

For a discrete image $f_{i,j}$ with $i \in 1, \ldots, N_1$ and $j \in 1, \ldots, N_2$, where N_1 and N_2 are the image dimensions, one defines *moments* of order $p + q$ as follows [14]:

$$m_{p,q} = \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} i^p \, j^q \, f_{i,j} \tag{4.1}$$

with $p = 0, \ldots, N_1 - 1$ and $q = 0, \ldots, N_2 - 1$

29

As one can see, the moments are defined for the whole image, meaning that for the computation of the moments of order $0, \ldots, N_1 + N_2 - 2$ we consider the values of all the image pixels at once.

These $N_1 N_2$ moments are sufficient to uniquely determine the discrete image $f_{i,j}$!

4.2   Central Moments

As we have mentioned already, in some applications it is useful for the moments to be invariant under translational motion. We can achieve this invariance by modifying the raw moments and computing new types of moments, namely *central moments*, which are invariant under the desired type of motion.

The *central moments* for a discrete image $f_{i,j}$ are defined as follows [14]:

$$\mu_{p,q} = \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} (i - \bar{i})^p \, (j - \bar{j})^q \, f_{i,j} \tag{4.2}$$

where $\bar{i} = \frac{m_{1,0}}{m_{0,0}}$ and $\bar{j} = \frac{m_{0,1}}{m_{0,0}}$ and they denote the coordinates of the center of gravity.

If we take a closer look at the above equation, we will see that it differs from the equation for the raw moments (4.1), only in that, instead of $i^p \, j^q$, in (4.2) we have $(i - \bar{i})^p \, (j - \bar{j})^q$. This difference determines the translation invariance of the central moments, as they have been "moved" in the center of the region, which means that no matter where the object is situated, the moments will be computed as if the object were in the center of the region.

4.3   Normalized Central Moments

After we have the translation invariant moments, we would also like to have moments that are invariant under (spatial) rescaling. This is also important in object recognition applications, as we may want to compare two similar objects, which differ only in their scale.

We can add scale invariance to the central moments, by normalizing them [14]. And so we have the following *normalized central moments*:

$$\eta_{p,q} = \frac{\mu_{p,q}}{\mu_{0,0}^{\gamma}} \tag{4.3}$$

with $\gamma = \frac{p + q}{2} + 1$ and $p + q = 2, 3, \ldots$

4.4   Rotation Invariant Moments

We can go even further and require that our moments are invariant under translational, scaling AND rotational transformations. In 1962, Hu derived his famous seven rotation invariant moments [18].

These are the seven *rotation invariant moments* that Hu described:

$$
\begin{aligned}
\phi_1 &= \eta_{2,0} + \eta_{0,2} \\
\phi_2 &= (\eta_{2,0} - \eta_{0,2})^2 + 4\eta_{1,1}^2 \\
\phi_3 &= (\eta_{3,0} - 3\eta_{1,2})^2 + (3\eta_{2,1} - \eta_{0,3})^2 \\
\phi_4 &= (\eta_{3,0} + \eta_{1,2})^2 + (\eta_{2,1} + \eta_{0.3})^2 \\
\phi_5 &= (\eta_{3,0} - 3\eta_{1,2})(\eta_{3,0} + \eta_{1,2})[(\eta_{0,3} + \eta_{1,2})^2 - 3(\eta_{2,1} + \eta_{0,3})^2] + \\
&\quad (3\eta_{2,1} - \eta_{0,3})(\eta_{2,1} + \eta_{0,3})[3(\eta_{3,0} + \eta_{1,2})^2 - (\eta_{2,1} + \eta_{0,3})^2] \\
\phi_6 &= (\eta_{2,0} - \eta_{0,2})[(\eta_{3,0} + \eta_{1,2})^2 - (\eta_{2,1} + \eta_{0,3})^2] + \\
&\quad 4\eta_{1,1}(\eta_{3,0} + \eta_{1,2})(\eta_{2,1} + \eta_{0,3}) \\
\phi_7 &= (3\eta_{1,2} - \eta_{3,0})(\eta_{3,0} + \eta_{1,2})[(\eta_{3,0} + \eta_{1,2})^2 - 3(\eta_{2,1} + \eta_{0,3})^2] + \\
&\quad (3\eta_{2,1} - \eta_{0,3})(\eta_{2,1} + \eta_{0,3})[3(\eta_{3,0} + \eta_{1,2})^2 - (\eta_{2,1} + \eta_{0,3})^2]
\end{aligned}
\tag{4.4}
$$

4.5 Moments for Colour Images

The above moments definitions are valid for black-and-white images, where the value of $f_{i,j}$ is the grey value of the pixel with coordinates (i, j).

As the reader has already seen, the definition of the raw moments is the one on which all the subsequent definitions are based. So, if we modify the raw moments definition in such a way as to make it valid for colour images, the rest of the definitions will follow.

When we work with colour images, we actually work with the RGB values of the pixels. So we have to plug the R, G and B values into the moments definition. Having this in mind we will define the *generalized colour moments* [29] for a discrete colour image as follows:

$$
M_{p,q}^{abc} = \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} i^p \, j^q \, [R(i,j)]^a \, [G(i,j)]^b \, [B(i,j)]^c
\tag{4.5}
$$

with $p = 0, \ldots, N_1 - 1$ and $q = 0, \ldots, N_2 - 1$.

$M_{p,q}^{abc}$ is said to be a generalized colour moment of order $p + q$ and degree $a + b + c$.

In this work we consider the R, G and B channels separately and hence we will consider only generalized moments of degree 1, where $(a, b, c) \in \{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$.

4.6 Skipping translational invariance

So far we have presented the definitions of the scaling and rotational invariant moments, definitions which are based on the raw moments and the centralized moments. When the moments are used for object recognition it makes sense to incorporate all kinds

of transformation invariance in order to neutralize every possible deformation and displacement difference. When it comes to computation of optic flow though, we have different requirements. In optic flow we are trying to detect the movement that is introduced in the image, not neutralize it. So, can we still use the moments for optic flow computation? The answer is "yes", but we have to make some slight modifications.

Actually the only unwanted invariance is the translation invariance (as we are trying to detect and compute translational movements). We still want to have scaling invariance and rotational invariance.

Having the generalized colour moments (4.5), we can define the scaling moments as follows:

$$H_{p,q} = \frac{M_{p,q}}{M_{0,0}^{\gamma}} \tag{4.6}$$

with $\gamma = \dfrac{p+q}{2} + 1$ and $p + q = 2, 3, \ldots$

We also want to have rotation invariance, which again leads us to Hu's seven moments. The difference this time is that they should be based not on the $\eta_{p,q}$ scaling moments, but on the $H_{p,q}$ moments. So, the definition of the new rotationally invariant moments Φ_1, Φ_2, Φ_3, Φ_4, Φ_5, Φ_6 and Φ_7 is the same as the one in equations (4.4), with the only difference that we substitute the ηs with Hs:

$$
\begin{aligned}
\Phi_1 &= H_{2,0} + H_{0,2} \tag{4.7}\\
\Phi_2 &= (H_{2,0} - H_{0,2})^2 + 4H_{1,1}^2 \\
\Phi_3 &= (H_{3,0} - 3H_{1,2})^2 + (3H_{2,1} - H_{0,3})^2 \\
\Phi_4 &= (H_{3,0} + H_{1,2})^2 + (H_{2,1} + H_{0.3})^2 \\
\Phi_5 &= (H_{3,0} - 3H_{1,2})(H_{3,0} + H_{1,2})[(H_{0,3} + H_{1,2})^2 - 3(H_{2,1} + H_{0,3})^2] + \\
&\quad (3H_{2,1} - H_{0,3})(H_{2,1} + H_{0,3})[3(H_{3,0} + H_{1,2})^2 - (H_{2,1} + H_{0,3})^2] \\
\Phi_6 &= (H_{2,0} - H_{0,2})[(H_{3,0} + H_{1,2})^2 - (H_{2,1} + H_{0,3})^2] + \\
&\quad 4H_{1,1}(H_{3,0} + H_{1,2})(H_{2,1} + H_{0,3}) \\
\Phi_7 &= (3H_{1,2} - H_{3,0})(H_{3,0} + H_{1,2})[(H_{3,0} + H_{1,2})^2 - 3(H_{2,1} + H_{0,3})^2] + \\
&\quad (3H_{2,1} - H_{0,3})(H_{2,1} + H_{0,3})[3(H_{3,0} + H_{1,2})^2 - (H_{2,1} + H_{0,3})^2]
\end{aligned}
$$

We have also developed an additional moment, which we call the 0th rotational moment, which is as follows:

$$\Phi_0 = H_{1,0}^2 + H_{0,1}^2 \tag{4.8}$$

The formal proof that the presented eight rotationally invariant moments Φ_0, Φ_1, Φ_2, Φ_3, Φ_4, Φ_5, Φ_6 and Φ_7 are really rotationally invariant can be found in *Appendix A*.

4.7 Moment Invariants

We have derived eight rotationally, but not translationally, invariant moments. These moments we can convert to constancy assumptions, which can be used for optic flow

computation.

The moments are defined based on the pixels in the whole image. What we want though, is to split the image into regions and for every such region to compute the moments. Having this in mind we define our moments constancy assumptions as follows:

$$\Phi_i(x, y, t) = \Phi_i(x + u, y + v, t + 1) \tag{4.9}$$

meaning that we assume that in the selected region - the Φ_i^{th} moment for $i \in 0, ..., 7$ remains constant.

4.8   Summary

In this chapter we have presented the concept of moments and their translation, scaling and rotation invariant reformulations. We have also derived rotationally invariant moments, which are not based on the central moments. These newly derived moments can be used as constancy assumptions and we can compute the optic flow based on them.

In contrast to the photometric constancy assumptions, presented in the previous chapter, these moment constancy assumptions are not invariant under illumination changes. Nevertheless they represent a new type of invariance, namely scaling and rotational invariance and thus can be used for optic flow computation for sequences without illumination changes.

Chapter 5

Multichannel Variational Methods

In Chapter 2 we have introduced the variational methods, which are among the best known techniques for optic flow computation. The basic idea behind the variational methods is the minimization of an energy functional. This energy functional is composed out of two parts: a data term and a smoothness term. The data term puts constraints on the image data and assumes constancy of some image features. The smoothness term, respectively, puts constraints on the optic flow solution and assumes different types of smoothness for the final flow field.

We presented two variational methods: the method of Horn-Schunck [17] and the Brox *et al.* method [4]. As we have mentioned already - the Horn-Schunck method is based on the brightness constancy assumption, while the Brox *et al.* method is based on the brightness and gradient constancy assumptions and a more sophisticated smoothness term.

In Chapters 3 and 4 we have presented different types of constancy assumptions, namely photometric and moment invariance assumptions.

In this chapter we will show how we can modify the Horn-Schunck and the Brox *et al.* methods so that we can plug our new constancy assumptions in them. As we will demonstrate in Chapter 6, combining a good method with a suitable constancy assumption, leads to results which are better than all those previously reported in the literature.

5.1 The Motion Tensor Notation

The most widely used constancy assumption in the black-and-white images optic flow computation is the brightness constancy assumption. As we have mentioned before, if we are given a black-and-white image f, the brightness constancy assumption states that:

$$f(x, y, t) = f(x + u, y + v, t + 1)$$

After a Taylor expansion the above statement becomes:

$$f_x u + f_y v + f_t = 0$$

The constancy assumptions are usually penalized in a quadratic way, so the data term used in the Horn and Schunk method is actually:

$$(f_x u + f_y v + f_t)^2$$

which, taking into account the notation $w^T = (u, v, 1)$, can be further simplified to:

$$(f_x u + f_y v + f_t)^2 = (w^T \nabla f)^2 = w^T (\nabla f \nabla f^T) w = w^T J_0(\nabla f) w$$

The matrix $J_0(\nabla f) = \nabla f \nabla f^T$ contains all necessary information to describe the constraint on the local solution imposed by the grey-value assumption. This matrix is known in the literature as the *motion tensor* [6], which represents the above stated constancy assumption.

Here we have used the grey-value constancy assumption to illustrate the motion tensor. However, this notation is, of course, valid for every other assumption that we have.

If we have more than one constancy assumption and we want to combine them in the data term, the newly constructed data term will have the following representation:

$$\sum_{i=1}^{n} \gamma_i (f_{i_x} u + f_{i_y} v + f_{i_t}) \tag{5.1}$$

where γ_i are positive weights for the different f_i assumptions for $i \in 1, ..., n$.

The above equation can be also represented in the motion tensor notation:

$$\sum_{i=1}^{n} \gamma_i (f_{i_x} u + f_{i_y} v + f_{i_t})^2 = \sum_{i=1}^{n} \gamma_i (w^T \nabla f_i)^2 = \sum_{i=1}^{n} \gamma_i (w^T \nabla f_i \nabla f_i^T w) = \tag{5.2}$$

$$= w^T \Big(\sum_{i=1}^{n} \gamma_i \nabla f_i \nabla f_i^T \Big) w = w^T \Big(\sum_{i=1}^{n} \gamma_i J_i(\nabla f_i) \Big) w = w^T J(\nabla f_1, ..., \nabla f_n) w$$

where $J_i(\nabla f_i) = \nabla f_i \nabla f_i^T$ is the motion tensor for the ith constancy assumption and $J(\nabla f_1, ..., \nabla f_n)$ is the so called *general motion tensor* [6]. This notation allows us to implement the variational methods much easier in practicle applications.

5.2 Multichannel Horn-Schunk Method

As the reader remembers, the Horn-Schunck method was based on the grey-value constancy assumption and was represented by the following energy functional:

$$E(w) = \int_{\Omega} \Big((f_x u + f_y v + f_t)^2 + \alpha(|\nabla u|^2 + |\nabla v|^2) \Big) \, dx dy \tag{5.3}$$

This method can be modified for colour image sequences and for more than one constancy assumption. If we work with colour images, the constancy assumptions f_i will be based on the colour information (not the grey-value). If we want to combine more than one assumption, we can do that using the general motion tensor notation (5.2).

Having this in mind, we can modify the Horn-Schunck method in the following manner:

$$E(w) = \int_\Omega \left(w^T \Big(\sum_{i=1}^n \gamma_i J_i(\nabla f_i) \Big) w + \alpha(|\nabla u|^2 + |\nabla v|^2) \right) dx dy \qquad (5.4)$$

We call the above modified Horn-Schunck method the *multichannel Horn-Schunck method.*

5.3   Multichannel Brox *et al.* Method

The method of Brox *et al.* is much more complicated than the method of Horn-Schunck and as such gives much better results. The method uses the grey-value constancy assumption (5.5) and the gradient constancy assumption (5.6), which this time are penalized in a non-quadratic way using the (5.7) penalizer.

$$D_{brightness}(f, w) = |f(x + u, y + v, t + 1) - f(x, y, t)|^2 \qquad (5.5)$$

$$D_{gradient}(f, w) = |\nabla f(x + u, y + v, t + 1) - \nabla f(x, y, t)|^2 \qquad (5.6)$$

$$\Psi(s) = \sqrt{s^2 + \varepsilon^2} \qquad (5.7)$$

As this method is one of the best known in the literature, we would like to modify it to work for colour image sequences and with multiple constancy assumption.

A similar strategy to the one in the previous section can be applied, where we are working with the f_i colour constancy assumptions and we are substituting them as a sum into the data term of the energy functional. The difference here, though, is that we can not use the motion tensor notation directly into the data term, as the constancy assumptions are not linearized in the Brox *et al.* method. However we can still benefit from the general, and easy to implement and integrate, form of the notation later in the numerics scheme (please refer to [6] for a more detailed explanation).

The original Brox *et al.* energy functional [4]

$$E(w) = \int_\Omega \Psi\Big(D_{brightness} + \gamma D_{gradient}\Big) + \alpha \Psi\Big(|\nabla u|^2 + |\nabla v|^2\Big) dx\, dy \qquad (5.8)$$

can be modified into

$$E(w) = \int_\Omega \Psi\Big(\sum_{i=1}^n \gamma_i \Big(f_i(x + u, y + v, t + 1) - f_i(x, y, t) \Big)^2 \Big) + \alpha \Psi\Big(|\nabla u|^2 + |\nabla v|^2\Big) dx\, dy$$

where we are again applying the robust penalization function to the invariance assumptions, but this time we are doing this to the sum of all the n non-linearized assumptions.

We call the above modified Brox *et al.* method the *multichannel Brox* et al. *method*.

5.4 Summary

In this chapter we have showed how the famous methods of Horn-Schunck and Brox *et al.* can be transformed into multichannel methods. In this way we can benefit from the features of the methods and in the same time integrate them with multiple colour images constancy assumptions.

In the next chapter we will present our experiments, which are based on these modified multichannel methods and the constancy assumptions described in Chapters 3 and 4.

Chapter 6

Experimental Results

In this chapter we will present experimental results that illustrate the theory from the previous chapters. First, we will shortly discuss the error measures and the visualization techniques for the computed optic flow solution. Then, we will give a short description of the used test sequences and the actual program implementation.

In Chapter 2 we presented the best methods for optic flow computation, namely the variational methods. These methods are based on the minimization of an energy functional, composed of a data and a smoothness terms. In Chapter 3 we developed suitable constancy assumptions for the data term that impose invariance on the solution under severe illumination changes. In Chapter 4 we derived constancy assumptions that are invariant under scaling and rotational transformations. The second and the third part of this chapter give the experimental results that illustrate the theory from the above mentioned chapters.

So far, while presenting the variational optic flow methods, we have been illustrating the theory first based on the simple Horn-Schunck method and then extending it to the much more sophisticated method of Brox *et al.* In this chapter we will again follow the same strategy and present the results first for the Horn-Schunck and then for the Brox *et al.* method.

6.1 Experimental Setup

Before presenting the actual experiments and results we will discuss briefly how we are going to assess the errors, the way we are going to visualize the solutions and the test sequences we have used in our experiments.

6.1.1 Error Measures

As we are about to perform some experiments we need a way to measure how good or bad they are. As mentioned before, the optic flow field is a vector field (Figure 6.1 - Right). For every pixel we have a vector that starts from the original position of the pixel and ends in the new position of the pixel.
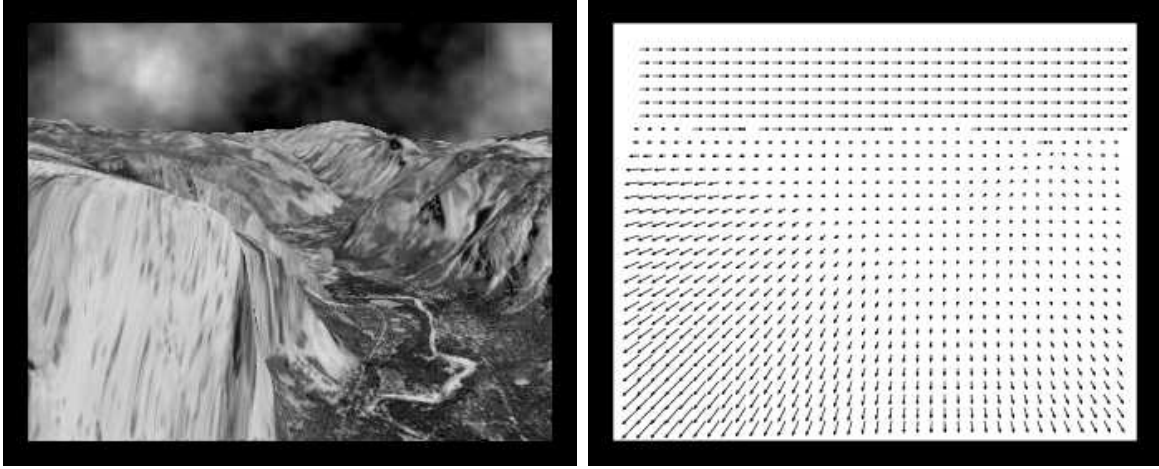
Figure 6.1: **Left:** The 9th frame of the Yosemite sequence; **Right:** The ground truth vector field, estimated between the 8th and the 9th frames of the same sequence.

For our experiments we are mainly using computer generated image sequences and for such sequences the ideal flow field is known. This flow field is also called *ground truth*. Having the ideal solution available, we can compare our results with it. Our solution is a vector field, so we need a way to compare two vector fields. As the important properties of a vector are its direction and magnitude, for each element in the vector field we will compare (see Figure 6.2):

1. the angle between every two vectors (the true one and the computed one)

2. the difference in the angles' magnitude

Let us consider that the correct flow field is denoted by $w_c = (u_c, v_c, 1)^T$ and the one that we have estimated is denoted by $w_e = (u_e, v_e, 1)^T$. We define one of the most widely used error measures - the so called **average angular error** (AAE) [1], as:

$$AAE(w_c, w_e) = \frac{1}{|\Omega|} \int_{\Omega} arccos\Big(\frac{w_e^T w_c}{|w_e||w_c|}\Big) dx \qquad (6.1)$$

where $|\Omega| = \int_{\Omega} dx$

To make it more clear we can expand the formula a bit more:

$$arccos\Big(\frac{w_e^T w_c}{|w_e||w_c|}\Big) = arccos\Big(\frac{u_e u_c + v_e v_c + 1}{\sqrt{(u_e^2 + v_e^2 + 1)(u_c^2 + v_c^2 + 1)}}\Big) \qquad (6.2)$$

The name of the error (*average angular error*) comes from the *averaging* over the whole image domain and the computation of the *angles* between the vectors.
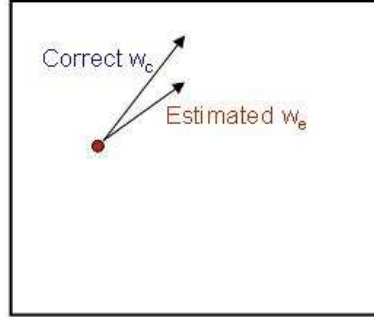
Figure 6.2: The difference between the correct vector field and the estimated vector field. The red dot represents a pixel from the image.

As we work with discrete images, the above formula has to be reformulated as follows:

$$AAE(w_c, w_e) = \frac{1}{|\Omega|} \sum_{\Omega} arccos\Big(\frac{w_e^T w_c}{|w_e||w_c|}\Big) \tag{6.3}$$

where Ω is the $N_1 \times N_2$ discrete image domain.

The AAE is not the only available error measurement in the literature. There are also other types of error measurements, like the ASE (average squared error) [16]. However, we have decided to use the AAE for estimating our results, because it is the most widely used one and most of the available results in the literature are reported based on it.

6.1.2 Optic Flow Visualization

We have already presented an error measurement technique for estimating our solution. Of course, this is the most precise way to measure the quality of the solution. However, the average angular error can be computed only for synthetic sequences, for which we have the ground truth. Obviously, when we want to compute the flow field for a real-world sequence we will need a different way to evaluate the computed flow field. Here follow some visualization techniques we can use.

Vector plot   One way to represent the solution is through a vector plot of the flow field (Figure 6.3 - Left). This way of representation shows the vectors themselves and thus we can see their direction and magnitude. However, this plot is sparse (because if we draw the vector for every pixel, we will not be able to see anything) and also not that easy to comprehend.

Colour plot   Because of the above stated problems of the vector plot representation, a colour representation has been developed (see Figure 6.3 - Right). We use the following
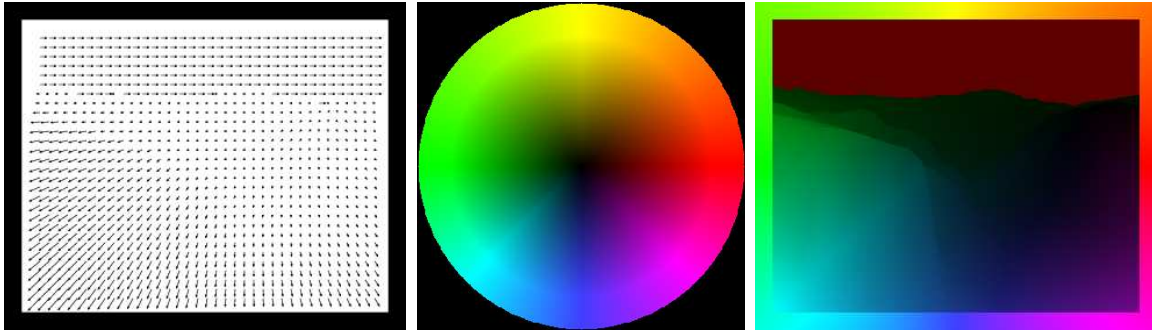
Figure 6.3: **Left:** Vector plot of the optic flow field solution; **Center:** Colour wheel, used for the colour coding; **Right:** Colour plot representation of the flow field.

coding for denoting the direction of the vector: if the vector points to the right, it is coded red; if it points to the left, it is coded green; if it points upwards, it is coded yellow and if it points downwards, it is coded blue. All vectors that point in some other mid-directions are coded accord to the colour wheel (Figure 6.3 - Center).

For denoting the magnitude of the vectors, we use the brightness of the colour - the greater the vector's magnitude is, the brighter the colour is. The resulting colour coded optic flow is dense and much easier to comprehend. This is why we are going to use this type of representation in this chapter.

6.1.3   Test Sequences

The main goal of this thesis is the computation of the optic flow field for colour image sequences. As this is a relatively young research field, there are not that many suitable test sequences available.

However, there are several sequences, that are well-known and there is a ground truth computed for them. In the following section we will give a brief introduction to the test sequences that we have chosen to use for our experiments. These sequences feature various types of motion, outliers and discontinuities. First, we will present the sequences that we use for our photometric experiments and then the sequences used for the rotational experiments.

Yosemite sequence with Clouds: This is one of the most widely used black-and-white sequences (see Figure 6.4). As we have been using this sequence to illustrate some of the optic flow principles throughout the chapters, we decided to present it here, even though we will not use it in our experiments.

The Yosemite sequence is a synthetic grey-value sequence that introduces both divergent and translational motion. The camera is moving backwards from the scene, while the clouds in the sky are moving to the right. The sequence contains 15 frames in total with maximum 5 pixels displacement between every two of the frames and can be downloaded from `ftp://ftp.csd.uwo.ca/pub/vision`.
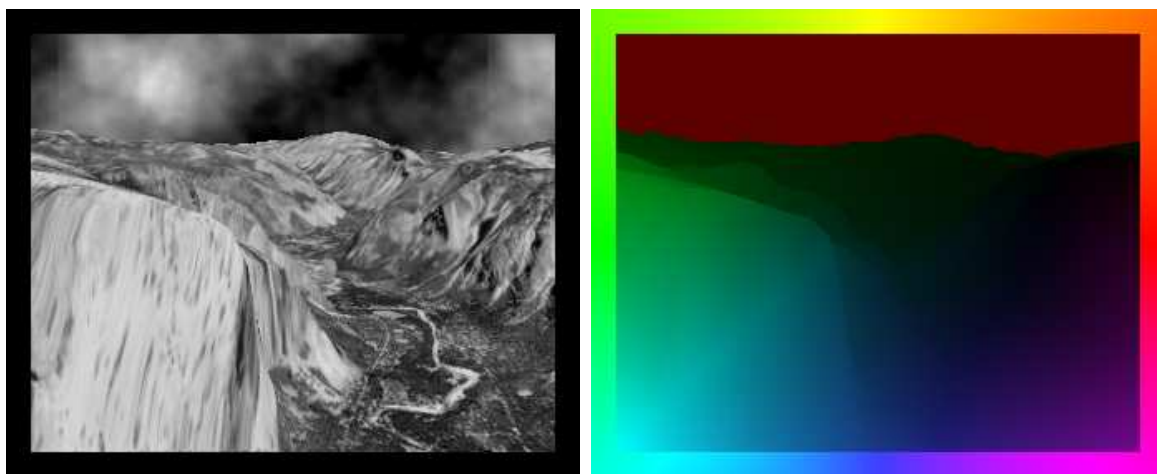
Figure 6.4: **Left:** Frame 8 of the Yosemite sequence; **Right:** Colour plot of the ground truth between the 8th and the 9th frames of the sequence.
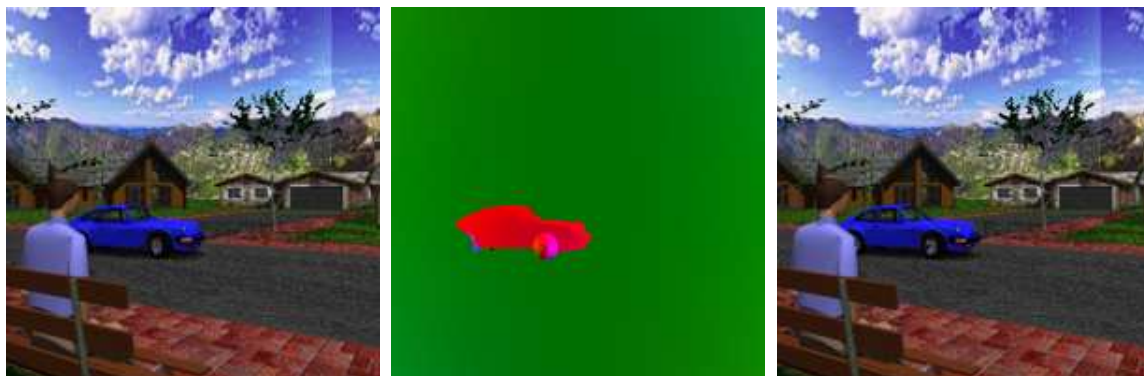

Figure 6.5: **Left:** Frame 10 of the "Street" sequence; **Center:** Colour plot of the ground truth between the 10th and the 11th frames; **Right:** Frame 11.

Street sequence: This is a synthetic colour image sequence that illustrates translational motion in front of a moving background (see Figure 6.5). As this is a synthesized sequence we have the ground truth available for comparison. The main motion here is as follows: the camera is moving to the right (which means that the background is moving to the left) and, at the same time, the car is moving to the right with a speed greater than the speed of the camera. This can be easily seen from the colour plot of the flow field (see Figure 6.5 - Center), where everything in the background is coloured green (as green denotes displacement to the left) and the car is coded in red (as red denotes displacement to the right). The largest displacements introduced are 4.5 pixels. This sequence contains 150 frames in total and can be downloaded from the following web-page `http://of-eval.sourceforge.net/`. The frames are square 200×200 images and we randomly selected the 10th and the 11th frames for our experiments. For more details regarding the sequence, please refer to [26].

Figure 6.6: **Left:** Frame 10 of the "Office" sequence; **Center:** The colour plot of the ground truth between the 10th and the 11th frames; **Right:** Frame 11.

Office sequence: This is another synthetic colour image sequence, where a different type of motion, namely divergent one, outliers and discontinuities are introduced (see Figure 6.6). We again have the ground truth available, which clearly illustrates the outward motion of the pixels as the camera is moving backwards from the scene. The maximum displacement between the frames is 1.5 pixels.

This sequence contains 60 frames in total and all of the 200×200 images can be downloaded from `http://of-eval.sourceforge.net/`. We again randomly selected the 10th and the 11th frames for our experiments.

DIPLODOC Road Stereo sequence: This is a real-world sequence, introducing divergent type of motion (see Figure 6.7). As this is a real-world sequence, we do not have ground truth available, so we can not compare the results, based on the average angular error. We can only judge our results based on the *computed* flow field, which we represent colour coded.

The sequence is a stereo one, taken with two different cameras, a left and a right camera, mounted on a moving vehicle. There are 865 320×240 image pairs available and we chose the 205th and the 207th left frames for our experiments. The sequence itself can be downloaded from `http://tev.itc.it/DATABASES/road.html`.

Slowly Rotating Texture sequence: The Slowly Rotating Texture sequence is again a synthetic colour sequence (see Figure 6.8) [6]. The sequence is also available in its black-and-white variant. There are 10 frames in total in the sequence and each frame is a 127×127 image.

The introduced motion pattern is rotation to the left on top of a static background. This can be easily seen from the colour-coded ground truth. The angle of rotation between two consecutive frames of the sequence is 5^{0} degrees. Thus, the displacement is relatively small, at least at locations that are situated closer to the center of rotation.

Figure 6.7: **Left:** Frame 205 of the Road sequence. There are illumination changes introduced in this frame due to the camera automatic adjustments; **Right:** Frame 207 of the same sequence.

Figure 6.8: **Left:** Frame 1 of the Slowly Rotating Texture sequence; **Center:** The colour plot of the ground truth between the 1st and the 2nd frames of the same sequence; **Right:** Frame 2 of the Slowly Rotating Texture sequence.

Figure 6.9: **Left:** Frame 10 of the Rotating Sphere sequence; **Center:** The colour plot of the ground truth between the 10th and the 11th frames of the same sequence; **Right:** Frame 11 of the Rotating Sphere sequence.

.

Rotating Sphere sequence   The last test sequence that we will use in our experiments is the so called Rotating Sphere sequence (see Figure 6.9). It contains 20 frames of size 200×200 and shows a rotating sphere. The sphere is rotating around its own axis in direction left-to-right on top of a stationary background.

The sequence, as well as its ground truth, are available for download from the following web-page: `http://of-eval.sourceforge.net/`. As one can see from Figure 6.9, the direction of the movement is again easy to detect from the colour plot of the flow field. The background is black, because it remains static in the sequence and only the sphere is coloured.

6.1.4   Implementation Details

After we have presented the test sequences that we are going to use for our experiments, we should also present the current implementation used for computing the reported results.

The test program is written in ANSI C based on routines that perform the Horn-Schunck and the Brox *et al.* methods. The program is also equipped with routines that solve a linear/non-linear system of equations, using the SOR method. When a non-linear system has to be solved, 10 SOR inner iterations are used in a loop of a fixed point iteration scheme. Also, when the coarse-to-fine technique is applied, maximum 10 levels of warping are used [4].

Other modules that have also been used, were responsible for memory allocation, input and output of various image formats and visualization of the result in OpenGL dialogs.

All of the above mentioned routines were provided to the author by one of the co-authors of the Brox *et al.* method, namely Dr. Andrés Bruhn [6].

Around the above mentioned functionality was build a new one, that extends the Horn-Schunck and the Brox *et al.* methods to their multichannel variants.

The test program is also equipped with routines that compute each of the discussed photometric invariants, as well as routines that compute raw, scaled and rotational black-and-white and colour moments. There are also modules responsible for different ways of channels normalization and modules dealing with the setting of weights to the channels.

The rest of the new modules are used for visualization of the invariant channels and for assistance of the above mentioned routines.

The test program has a user-friendly menu, which allows the user to optimize the parameters of the methods in the course of the program execution and to monitor the performance of the methods with respect to the average angular error and the colour-coded flow field.

Concept	#Ch	Type	normal	mult.	mult.+add.
RGB	3	-	6.36^0	20.32^0	23.25^0
HSV (hue)	1	3	11.2^0	11.2^0	11.2^0
Spherical ($\phi\theta$)	2	2	7.08^0	7.08^0	8.64^0
Arithmetic mean	3	2	7.26^0	7.26^0	8.46^0
Geometric mean	3	2	7.96^0	7.96^0	12.74^0
$\nabla\ln(\mathbf{RGB})$	6	1	7.11^0	7.15^0	7.42^0

Table 6.1: *Comparison of the different illumination invariants for the "Street" sequence, using the Horn-Schunck multichannel method*

6.2 Photometric Experiments

In this section we are going to illustrate how exactly the photometric invariants, introduced in Chapter 3, work in practice. For this purpose we use the "Street" and the "Office" sequences with the Horn-Schunck and the Brox *et al.* methods.

We can judge the real practical value of the photometric invariants, only when some severe illumination changes appear, and that is why we have to use image sequences, where such changes exist. Unfortunately, so far there are no synthetic colour sequences available with different types of illumination changes introduced. The synthetic sequences allow us to compare our results with the ground truth and thus to evaluate how good or bad our experiments were. That is why we had to modify the available synthetic sequences in such a way as to introduce some spatially varying multiplicative and additive illumination changes.

At the end of this section, we will repeat the same experiments, but using the real-image sequence, where we do have realistic illumination changes. Thus we intend to prove that the photometric invariance assumptions that we have developed, are useful for real world data.

6.2.1 Street sequence experiment

Our first experiment is based on the "Street" sequence. We have the 10th and the 11th frames available and we have modified the 11th frame. The multiplicative illumination is represented by shading changes in the image, while the additive is represented by highlights. We have deliberately chosen the parts of the image to degrade, in particular the car, as this is the object that has to be detected. In our experiments we take frame 10 and frame 11, when we consider normal illumination, or we take frame 10 and one of the strongly degraded variants of frame 11, when we consider illumination changes.

The Horn-Schunck multichannel method   First, we will examine the photometric invariants, using the Horn-Schunck multichannel method (see Section 5.2). The results are presented in Table 6.1. In this table we have listed all of the eight pho-

tometric invariants, discussed in Chapter 3; the number of channels used during the computation; the type of the invariant (see Section 3.2.1) and the AAE for the three cases of illumination changes. All weights γ_i were set to 1 (see Chapter 5) and the rest of the parameters have been optimized with respect to the AAE.

As one can see from Table 6.1, the *RGB constancy assumption* gives the best results, among the rest of the assumptions, when we have normal illumination, but it completely fails under varying one. This, of course, is expected, because, in the case of normal illumination, we are using all the available information and do not discard any of it. However, when we introduce some changing illumination, the assumption is still depending on all of the information, but this information is misleading in these cases and thus it is no longer valid. This example proves the need of photometric invariants.

The *hue assumption*, as a Type 3 invariant, does not change at all (AAE 11.2^0), no matter how severe the illumination changes are. However, gaining invariance is at the expense of throwing away information, so the more invariant an assumption is, the more information is discarded. This is the case with the hue - it is invariant under all kinds of illumination, but the AAE is one of the worst.

In comparison, the other colour space assumption, $\phi\theta$, gives the best results among the rest of the assumptions with AAE of 7.08^0 degrees. We have considered the ϕ and the θ assumptions together, as they belong to the same type of assumptions (Type 2), the same range and the same colour space. We have tested every one of them separately and have concluded that they perform better when they are combined. This is one example how different assumptions can be combined and result in a better performance.

What also draws our attention is the performance of the *log-derivatives*, AAE 7.11^0, 7.15^0 and 7.42^0. Even though this assumption is not invariant under such illumination changes, it still performs relatively well and stable. This proves that constancy assumptions based on the derivatives principle [37, 4, 33] are always a good choice.

In conclusion, we could say that, the choice which photometric assumption to use is a trade-off between better precision (like in the case of the log-derivatives assumption) and stability under sever illumination changes (like in the case of the hue assumption).

The actual computed optic flow fields for *normal* illumination, *multiplicative* and *multiplicative and additive* illumination changes, can be seen in Figures 6.10, 6.11 and 6.12 respectively.

From the colour coded flow fields it again becomes obvious that the *RGB* assumption gives the worst results, when some illumination changes are introduced. Also, it is obvious that the *hue* assumption gives the same stable result in all three cases. Figures 6.11 and 6.12 nicely illustrate what happens with the flow field, when the assumption we are using is violated, like in the case of the *arithmetic mean* and *geometric mean*.

As it became clear from the AAE results, the *log-derivatives* assumption performs relatively stable in all the cases and this is obvious also from the actual colour-coded flow fields.
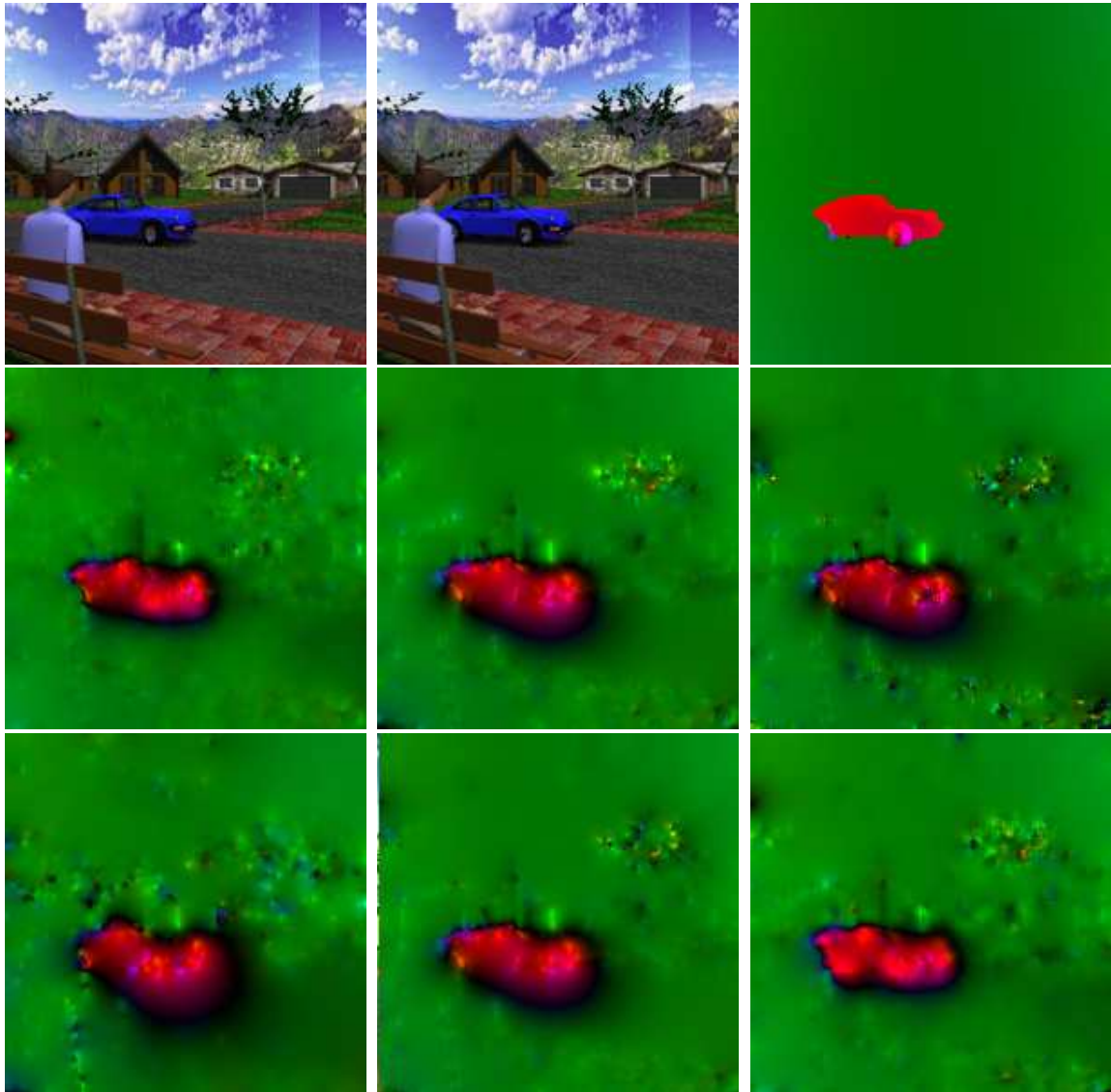
Figure 6.10: **Experiment 1.I:** No illumination changes **Top Left:** Frame 10 of the "Street" sequence; **Top Center:** Frame 11; **Top Right:** The ground truth colour-coded flow field between frame 10 and frame 11; **Middle Left:** The colour-coded flow field, computed with the Horn-Schunck multichannel method for the RGB constancy assumption; **Middle Center:** The arithmetic mean assumption; **Middle Right:** The geometric mean assumption; **Bottom Left:** The hue assumption; **Bottom Center:** The $\phi\theta$ assumption; **Bottom Right:** The log-derivatives assumption.

Figure 6.11: **Experiment 1.II:** Local multiplicative illumination changes **Top Left:** Frame 10 of the "Street" sequence; **Top Center:** Frame 11 (with multiplicative illumination changes); **Top Right:** The ground truth colour-coded flow field between frame 10 and frame 11; **Middle Left:** The colour-coded flow field, computed with the Horn-Schunck multichannel method for the RGB constancy assumption; **Middle Center:** The arithmetic mean assumption; **Middle Right:** The geometric mean assumption; **Bottom Left:** The hue assumption; **Bottom Center:** The $\phi\theta$ assumption; **Bottom Right:** The log-derivatives assumption.
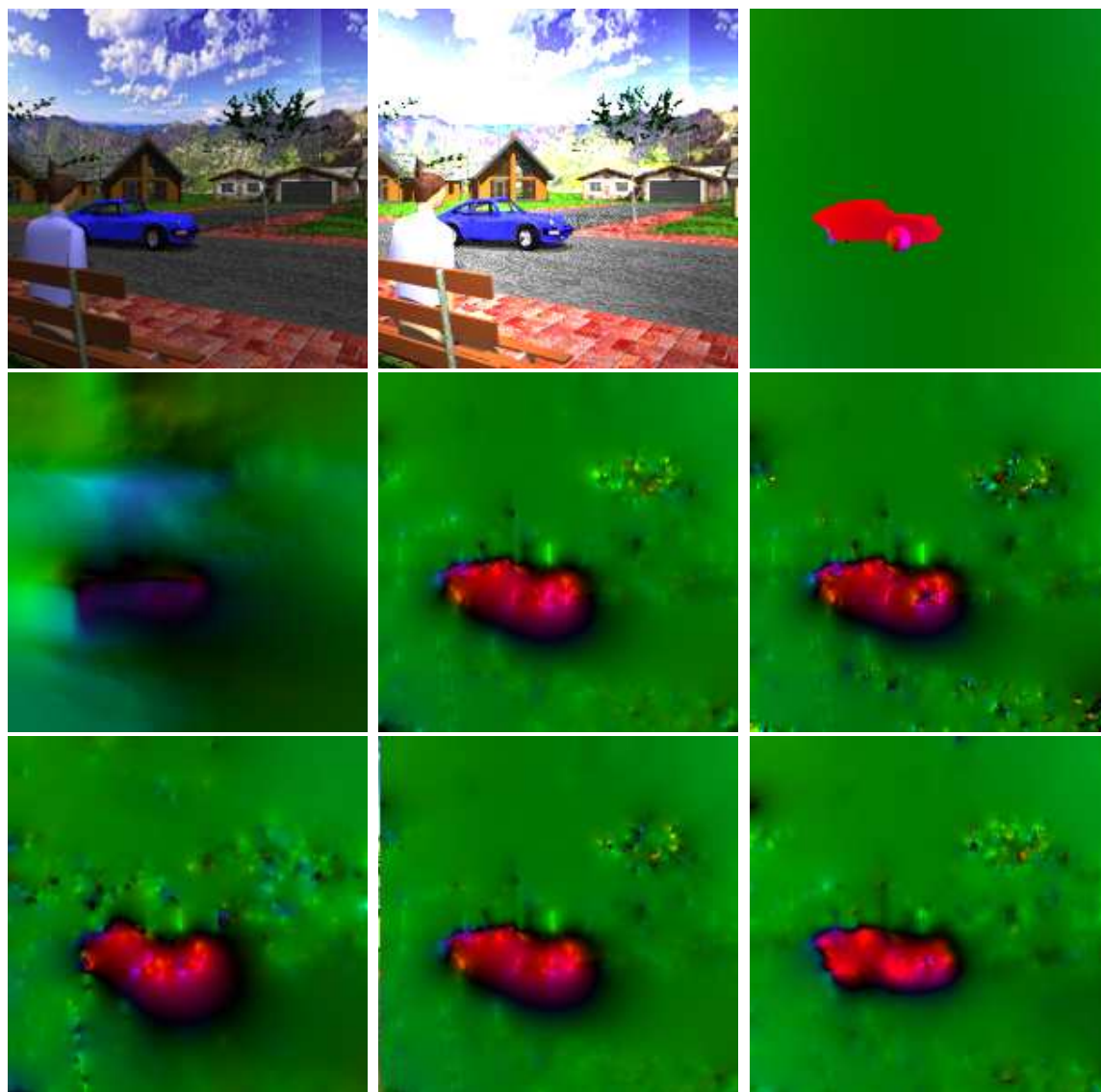
.
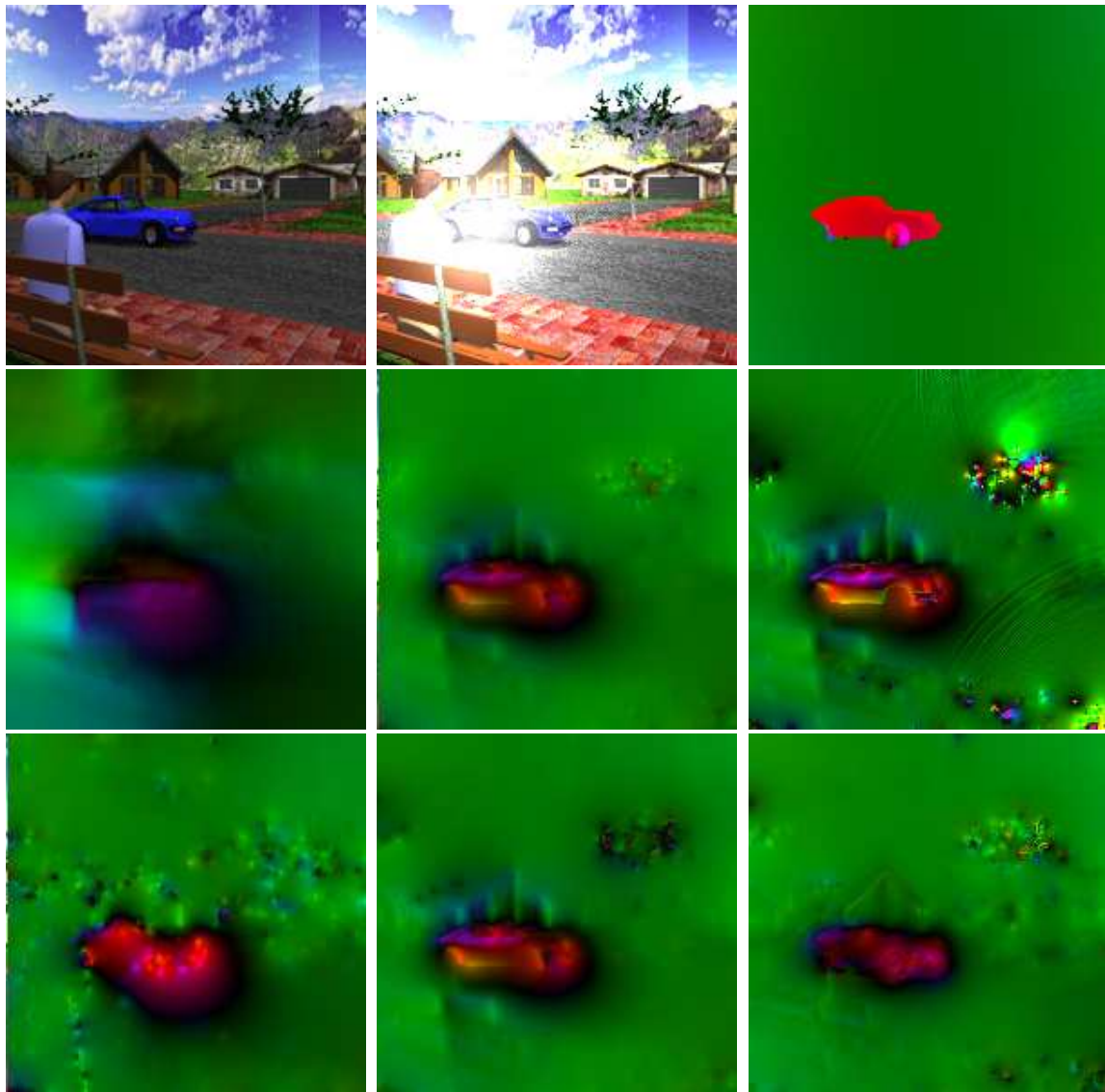
Figure 6.12: **Experiment 1.III:** Multiplicative and additive illumination changes **Top Left:** Frame 10 of the "Street" sequence; **Top Center:** Frame 11 (with multiplicative and additive illumination changes); **Top Right:** The ground truth colour-coded flow field between frame 10 and frame 11; **Middle Left:** The colour-coded flow field, computed with the Horn-Schunck multichannel method for the RGB constancy assumption; **Middle Center:** The arithmetic mean assumption; **Middle Right:** The geometric mean assumption; **Bottom Left:** The hue assumption; **Bottom Center:** The $\phi\theta$ assumption; **Bottom Right:** The log-derivatives assumption.

Concept	#Ch	Type	normal	mult.	mult.+add.
RGB	3	-	2.65^0	43.44^0	43.44^0
HSV (hue)	1	3	4.28^0	4.28^0	4.28^0
Spherical ($\phi\theta$)	2	2	2.07^0	2.07^0	3.37^0
Arithmetic mean	3	2	2.22^0	2.22^0	3.71^0
Geometric mean	3	2	2.26^0	2.26^0	5.64^0
$\nabla\ln(\mathbf{RGB})$	6	1	2.64^0	3.89^0	3.92^0

Table 6.2: *Comparison of the different illumination invariants for the "Street" sequence, using the Brox* et al. *multichannel method*

The Brox *et al.* multichannel method In our next experiment we perform the same tests like in the previous experiment, but this time we use the Brox *et al.* multichannel method (see Section 5.3).

The results are listed in Table 6.2. We again have set all the weights γ_i to 1 and we have optimized the rest of the parameters with respect to the AAE error.

As one can see from Table 6.2, the results this time are much better than the Horn-Schunck results in Table 6.1. Of course, this is due to the much more sophisticated approach that Brox *et al.* are applying.

We again see the same tendency in the performance of the photometric invariants. The standard *RGB assumption* fails completely under varying illumination. The *hue assumption* again remains unchanged, no matter what kind of illumination we have, and the *$\phi\theta$ assumption* again has the best performance among the other assumptions with AAE of 2.07^0 degrees.

For this experiment we have the result of 2.64^0 available from the original Brox *et al.* method for this sequence [6] and the result of 4.85^0 by Weickert and Schnör [40], which are the best results in the literature known to the author. These results are computed for the "Street" sequence with no varying illumination introduced. One can notice that the results that we have received in Table 6.2 outperform the best results available in the literature, even when we have severe illumination changes. This proves the usefulness of the colour photometric invariants and the multichannel methods in general.

Looking at the results, one may also notice that the bad results here are even worse than the bad results in Table 6.1 (the AAE for the *RGB* assumption). This is due to the warping technique used in the Brox *et al.* method, where the errors, if any, are accumulating.

The computed flow fields can be seen on Figures 6.13, 6.14 and 6.15. When we compare these flow fields with the ones that we had with the Horn-Schunck method, we notice that the Brox *et al.* method produces fields with well-defined object edges. This effect comes from the piecewise smoothness of the smoothness term. If we take our best result for this sequence, namely the $\phi\theta$ one, we notice that the background flow field has been accurately estimated and the shape of the car is well preserved.
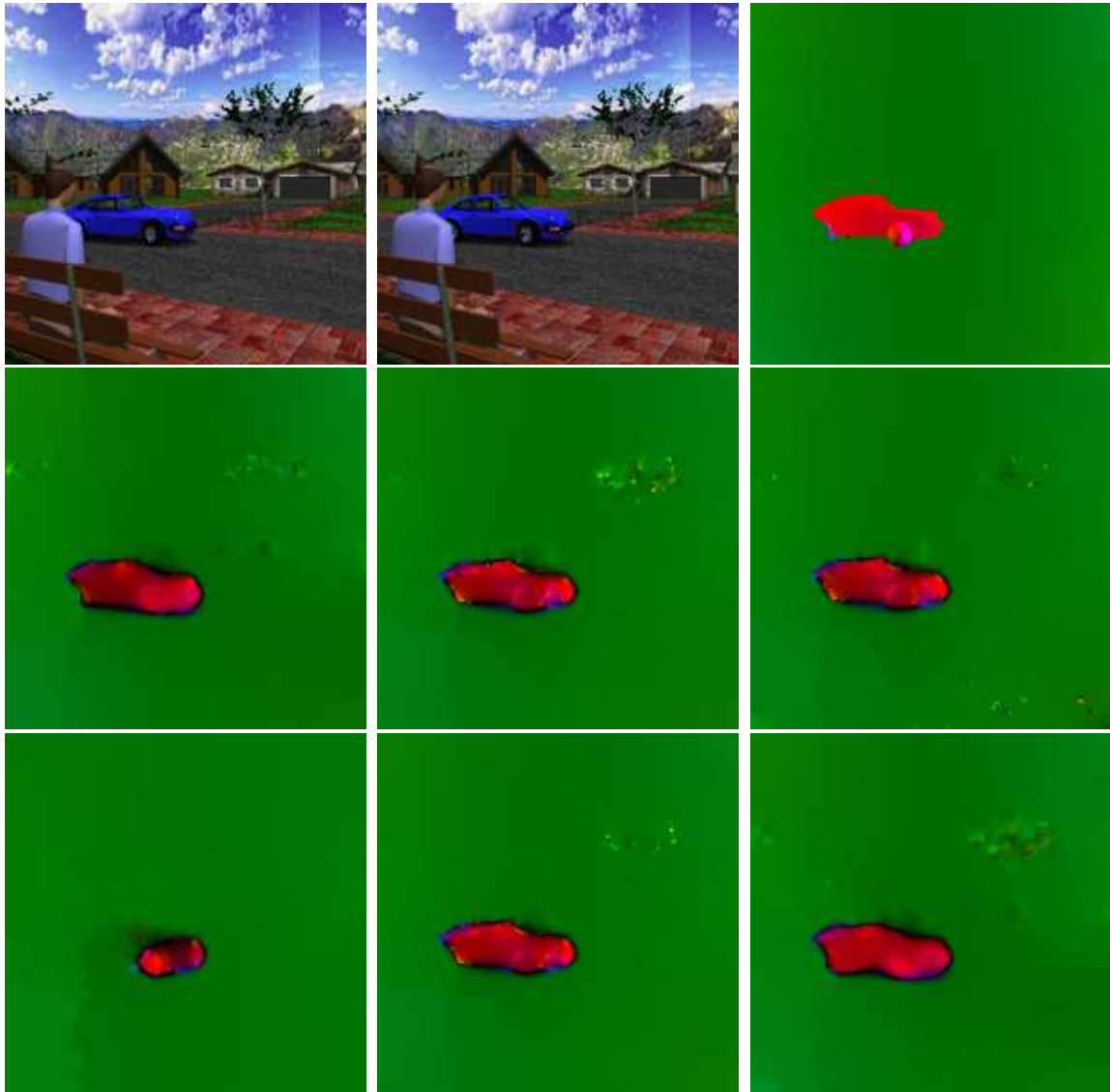
Figure 6.13: **Experiment 2.I:** No illumination changes **Top Left:** Frame 10 of the "Street" sequence; **Top Center:** Frame 11; **Top Right:** The ground truth colour-coded flow field between frame 10 and frame 11; **Middle Left:** The colour-coded flow field, computed with the Brox *et al.* multichannel method for the RGB constancy assumption; **Middle Center:** The arithmetic mean assumption; **Middle Right:** The geometric mean assumption; **Bottom Left:** The hue assumption; **Bottom Center:** The $\phi\theta$ assumption; **Bottom Right:** The log-derivatives assumption.

Figure 6.14: **Experiment 2.II:** Multiplicative illumination changes **Top Left:** Frame 10 of the "Street" sequence; **Top Center:** Frame 11 (with multiplicative illumination changes); **Top Right:** The ground truth colour-coded flow field between frame 10 and frame 11; **Middle Left:** The colour-coded flow field, computed with the Brox *et al.* multichannel method for the RGB constancy assumption; **Middle Center:** The arithmetic mean assumption; **Middle Right:** The geometric mean assumption; **Bottom Left:** The hue assumption; **Bottom Center:** The $\phi\theta$ assumption; **Bottom Right:** The log-derivatives assumption.
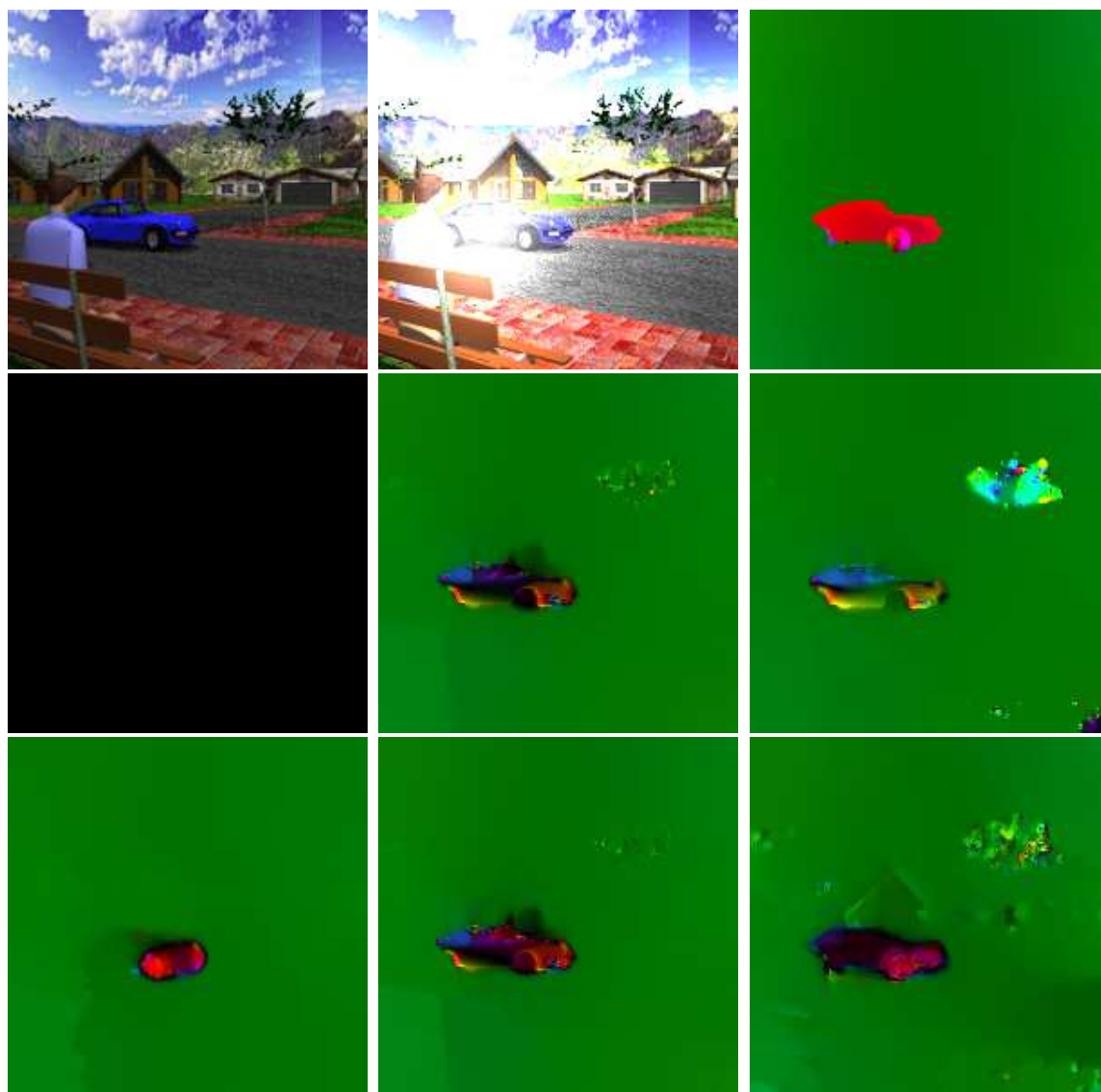
Figure 6.15: **Experiment 2.III:** Multiplicative and additive illumination changes **Top Left:** Frame 10 of the "Street" sequence; **Top Center:** Frame 11 (with multiplicative and additive illumination changes); **Top Right:** The ground truth colour-coded flow field between frame 10 and frame 11; **Middle Left:** The colour-coded flow field, computed with the Brox *et al.* multichannel method for the RGB constancy assumption; **Middle Center:** The arithmetic mean assumption; **Middle Right:** The geometric mean assumption; **Bottom Left:** The hue assumption; **Bottom Center:** The $\phi\theta$ assumption; **Bottom Right:** The log-derivatives assumption.

Concept	#Ch	Type	normal	mult.	mult.+add.
RGB	3	-	4.14^0	18.63^0	18.26^0
HSV (hue)	1	3	7.84^0	7.84^0	7.84^0
Spherical ($\phi\theta$)	2	2	7.29^0	7.29^0	10.71^0
Arithmetic mean	3	2	7.42^0	7.42^0	10.78^0
Geometric mean	3	2	8.64^0	8.64^0	12.92^0
$\nabla\ln(\mathbf{RGB})$	6	1	5.44^0	5.69^0	6.17^0

Table 6.3: *Comparison of the different illumination invariants for the "Office" sequence, using the Horn-Schunck multichannel method*

6.2.2   Office sequence experiment

With this second experiment we would like to cross-check the invariance of the photometric constancy assumptions. For this purpose we are going to use another image sequence, namely the "Office" sequence. In this sequence we have a different type of motion introduced - divergent. We will see that the invariance of our assumptions does not depend on the type of motion in the sequence and still remains constant. In this experiment frames 10 and 11 are used, and frame 11 is modified with spatially varying *multiplicative* and *multiplicative and additive* illumination.

Horn-Schunck multichannel method   We will once more start testing our photometric invariants with the Horn-Schunck multichannel method.

The average angular error results are listed in Table 6.3. Here we again consider all channels equally important and thus set the weights γ_i to 1. The parameters ρ and α of the variational method (see Section 2.1) are optimized with respect to the average angular error.

If we take a look at the AAE results, we will again see that every assumption is acting in accordance with its Type. The *RGB assumption* is again the best one in the normal illumination case, because it uses all the available colour information from the image, but fails in the rest of the cases. The *hue assumption* remains invariant, no matter what kind of illumination we introduce to the image. The *$\phi\theta$, arithmetic mean* and the *geometric mean* assumptions, as Type 2 assumptions, do not change under local multiplicative illumination. The *log-derivatives* assumption, even though not invariant in these cases and not completely stable, still performs well, and even better than the rest of the photometric invariants.

The actual computed colour-coded optic flow fields can be found on Figures 6.16, 6.17 and 6.18. The flow fields illustrate the same tendencies that we have observed from the AAE results. One can see how bad the *RGB* assumption performs under changing illumination conditions and how the flow field degrades for the photometric invariants under multiplicative and additive changes.
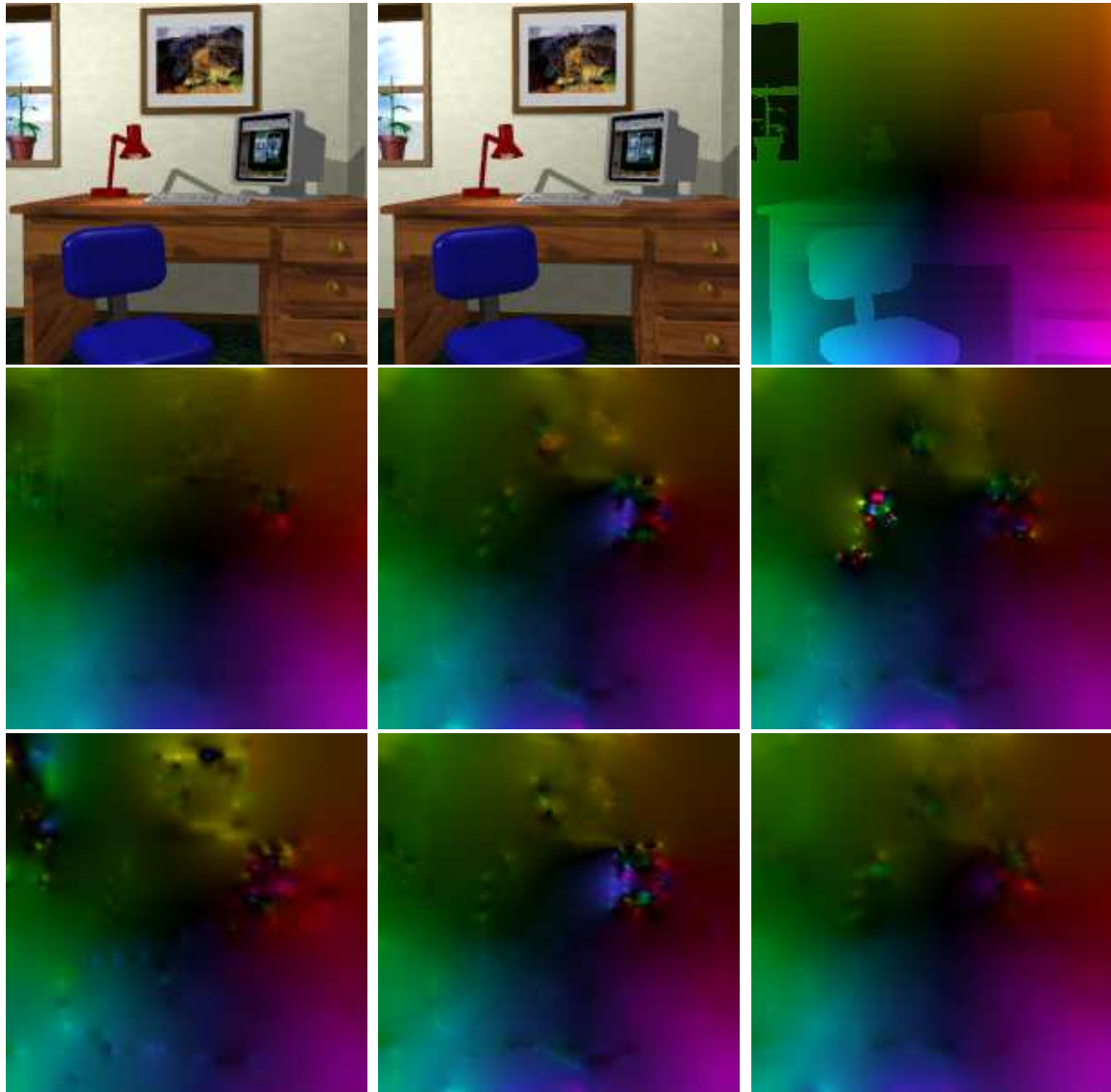
Figure 6.16: **Experiment 3.I:** No illumination changes **Top Left:** Frame 10 of the "Office" sequence; **Top Center:** Frame 11; **Top Right:** The ground truth colour-coded flow field between frame 10 and frame 11; **Middle Left:** The colour-coded flow field, computed with the Horn-Schunck multichannel method for the RGB constancy assumption; **Middle Center:** The arithmetic mean assumption; **Middle Right:** The geometric mean assumption; **Bottom Left:** The hue assumption; **Bottom Center:** The $\phi\theta$ assumption; **Bottom Right:** The log-derivatives assumption.
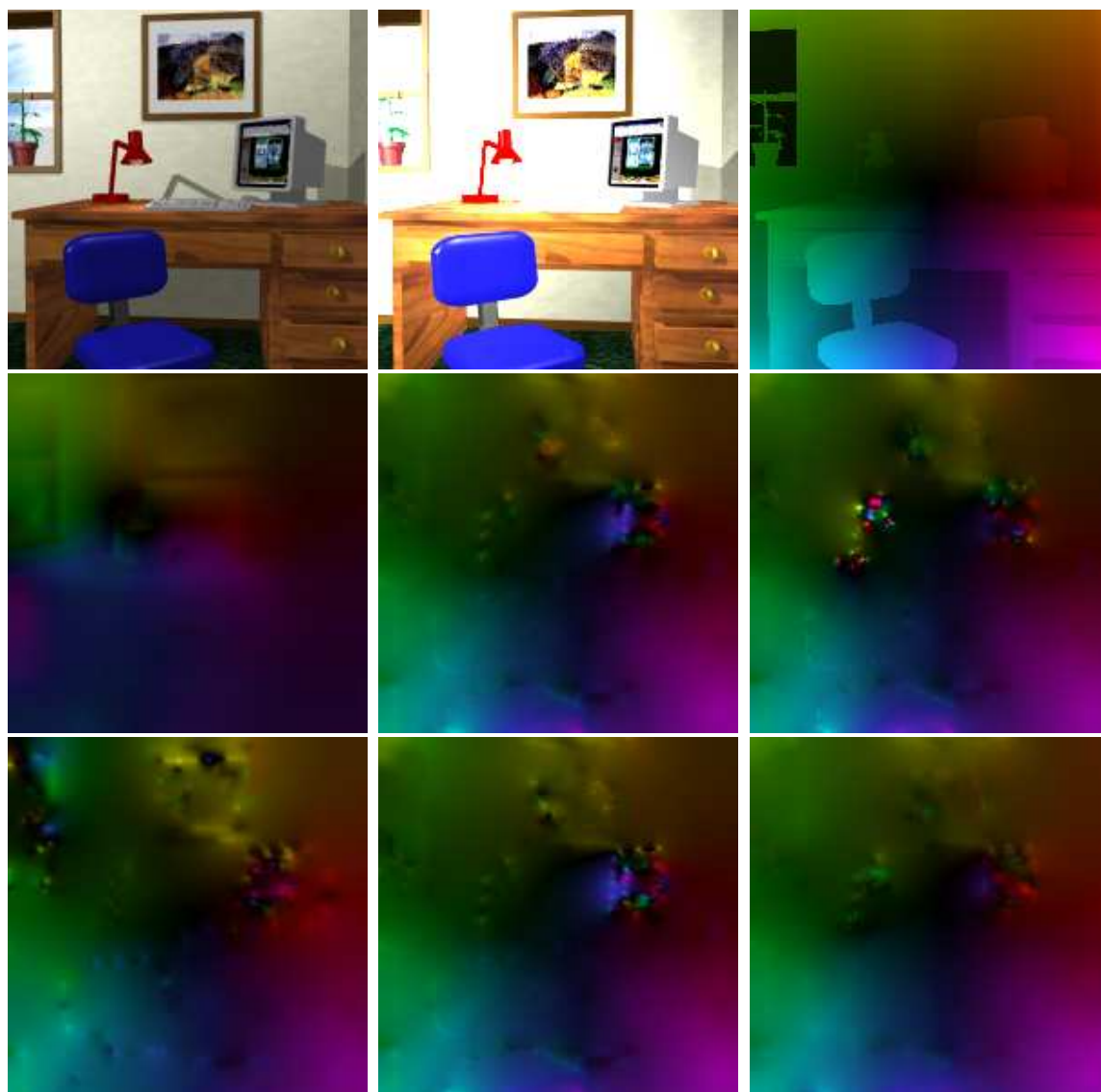
Figure 6.17: **Experiment 3.II:** Multiplicative illumination changes **Top Left:** Frame 10 of the "Office" sequence; **Top Center:** Frame 11 (with multiplicative illumination changes); **Top Right:** The ground truth colour-coded flow field between frame 10 and frame 11; **Middle Left:** The colour-coded flow field, computed with the Horn-Schunck multichannel method for the RGB constancy assumption; **Middle Center:** The arithmetic mean assumption; **Middle Right:** The geometric mean assumption; **Bottom Left:** The hue assumption; **Bottom Center:** The $\phi\theta$ assumption; **Bottom Right:** The log-derivatives assumption.
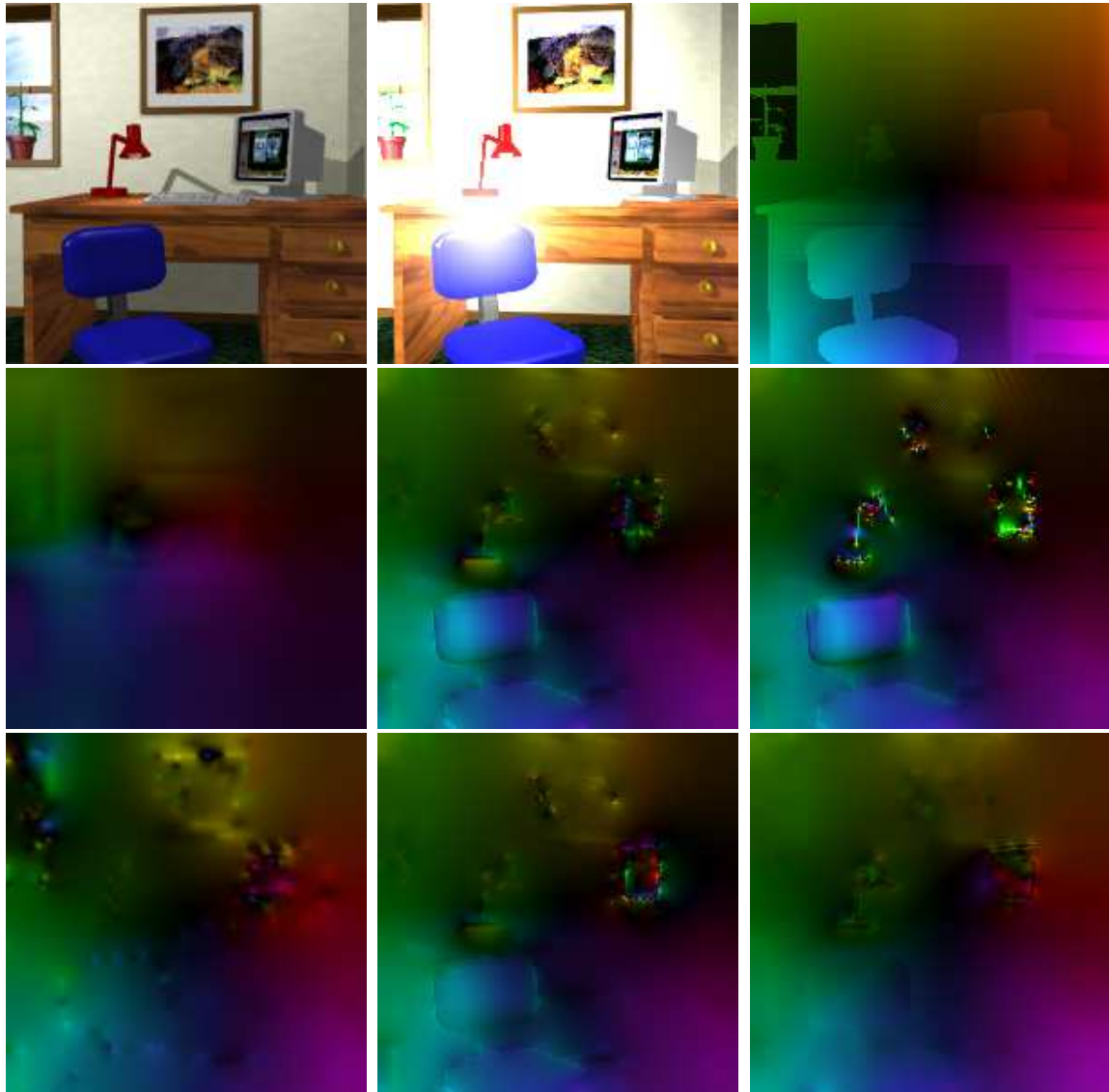
Figure 6.18: **Experiment 3.III:** Multiplicative and additive illumination changes **Top Left:** Frame 10 of the "Office" sequence; **Top Center:** Frame 11 (with multiplicative and additive illumination changes); **Top Right:** The ground truth colour-coded flow field between frame 10 and frame 11; **Middle Left:** The colour-coded flow field, computed with the Horn-Schunck multichannel method for the RGB constancy assumption; **Middle Center:** The arithmetic mean assumption; **Middle Right:** The geometric mean assumption; **Bottom Left:** The hue assumption; **Bottom Center:** The $\phi\theta$ assumption; **Bottom Right:** The log-derivatives assumption.

Concept	#Ch	Type	normal	mult.	mult.+add.
RGB	3	-	4.44^0	28.82^0	28.82^0
HSV (hue)	1	3	8.24^0	8.24^0	8.24^0
Spherical ($\phi\theta$)	2	2	6.82^0	6.82^0	8.52^0
Arithmetic mean	3	2	6.73^0	6.73^0	8.68^0
Geometric mean	3	2	7.27^0	7.27^0	9.01^0
$\nabla\ln(\mathbf{RGB})$	6	1	4.98^0	5.43^0	5.60^0

Table 6.4: *Comparison of the different illumination invariants for the "Office" sequence, using the Brox* et al. *multichannel method*

Brox *et al.* multichannel method   The tests that we presented in the previous paragraph will be repeated here, with the difference that we will use the Brox *et al.* multichannel method.

Table 6.4 presents the average angular error results for each of the discussed photometric invariants. The results again illustrate the specific invariance properties of the photometric constancy assumptions. Every assumption performs in accordance with its Type. Thus this experiment confirms once again the invariance of our photometric assumptions.

As the readers have probably noticed, the average angular errors in this case are almost the same as the errors that we had in the Horn-Schunck case (see Table 6.3). This is mainly due to the fact that the maximum pixel displacement between two consecutive frames of this sequence is 1.5, which does not give chance to the Brox *et al.* method to outperform Horn-Schunck by a high degree. For a comparison, the maximum displacement for the "Street" sequence is 5 pixels. The real power of the Brox *et al.* can be demonstrated for image sequences where we have large displacements. We should also note that here we are using the 2-*D* variant of the Brox *et al.*, where the time component is not taken into consideration. However, the edge preserving property of the method is still easily seen, when we compare the flow fields from Figures 6.19, 6.20 and 6.21 with the ones from Figures 6.16, 6.17 and 6.18.

With all of the above presented experiments we have demonstrated the invariance of the photometric assumptions. We have tested the assumptions against different types of motion, outliers and discontinuities and in all the cases they proved to be invariant. The results have been illustrated with the average angular error and the colour-coded flow fields. In conclusion, we would say, it definitely pays off to consider the usage of photometric invariance assumptions, when sequences with illumination changes are processed.

In the next part of this section we will demonstrate that the photometric invariants give good results for real-world data as well. This experiment is important, as we have deliberately developed the photometric assumptions to work for real-world illumination conditions.
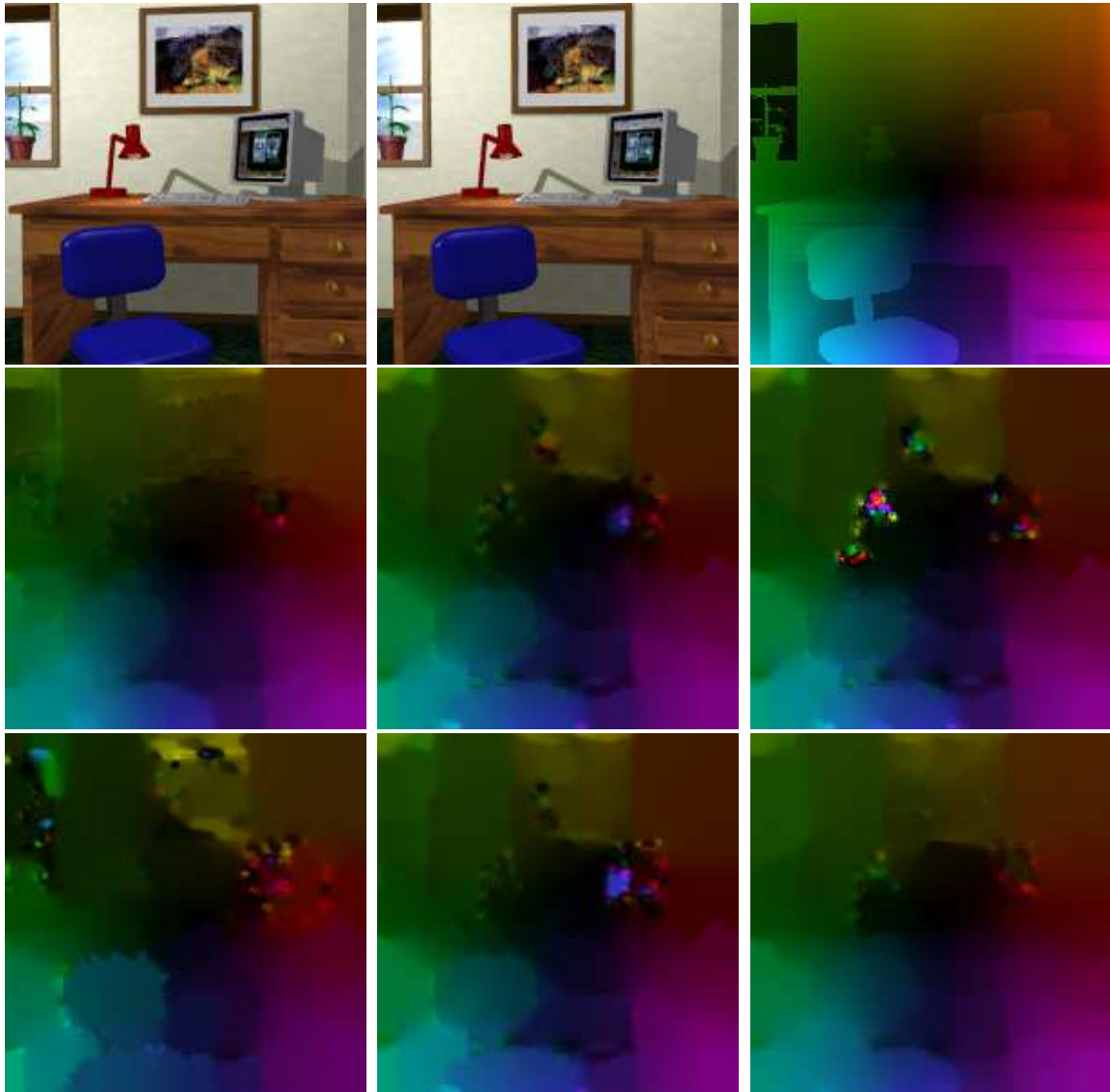
Figure 6.19: **Experiment 4.I:** No illumination changes **Top Left:** Frame 10 of the "Office" sequence; **Top Center:** Frame 11; **Top Right:** The ground truth colour-coded flow field between frame 10 and frame 11; **Middle Left:** The colour-coded flow field, computed with the Brox *et al.* multichannel method for the RGB constancy assumption; **Middle Center:** The arithmetic mean assumption; **Middle Right:** The geometric mean assumption; **Bottom Left:** The hue assumption; **Bottom Center:** The $\phi\theta$ assumption; **Bottom Right:** The log-derivatives assumption.

Figure 6.20: **Experiment 4.II:** Multiplicative illumination changes **Top Left:** Frame 10 of the "Office" sequence; **Top Center:** Frame 11 (with multiplicative illumination changes); **Top Right:** The ground truth colour-coded flow field between frame 10 and frame 11; **Middle Left:** The colour-coded flow field, computed with the Brox *et al.* multichannel method for the RGB constancy assumption; **Middle Center:** The arithmetic mean assumption; **Middle Right:** The geometric mean assumption; **Bottom Left:** The hue assumption; **Bottom Center:** The $\phi\theta$ assumption; **Bottom Right:** The log-derivatives assumption.
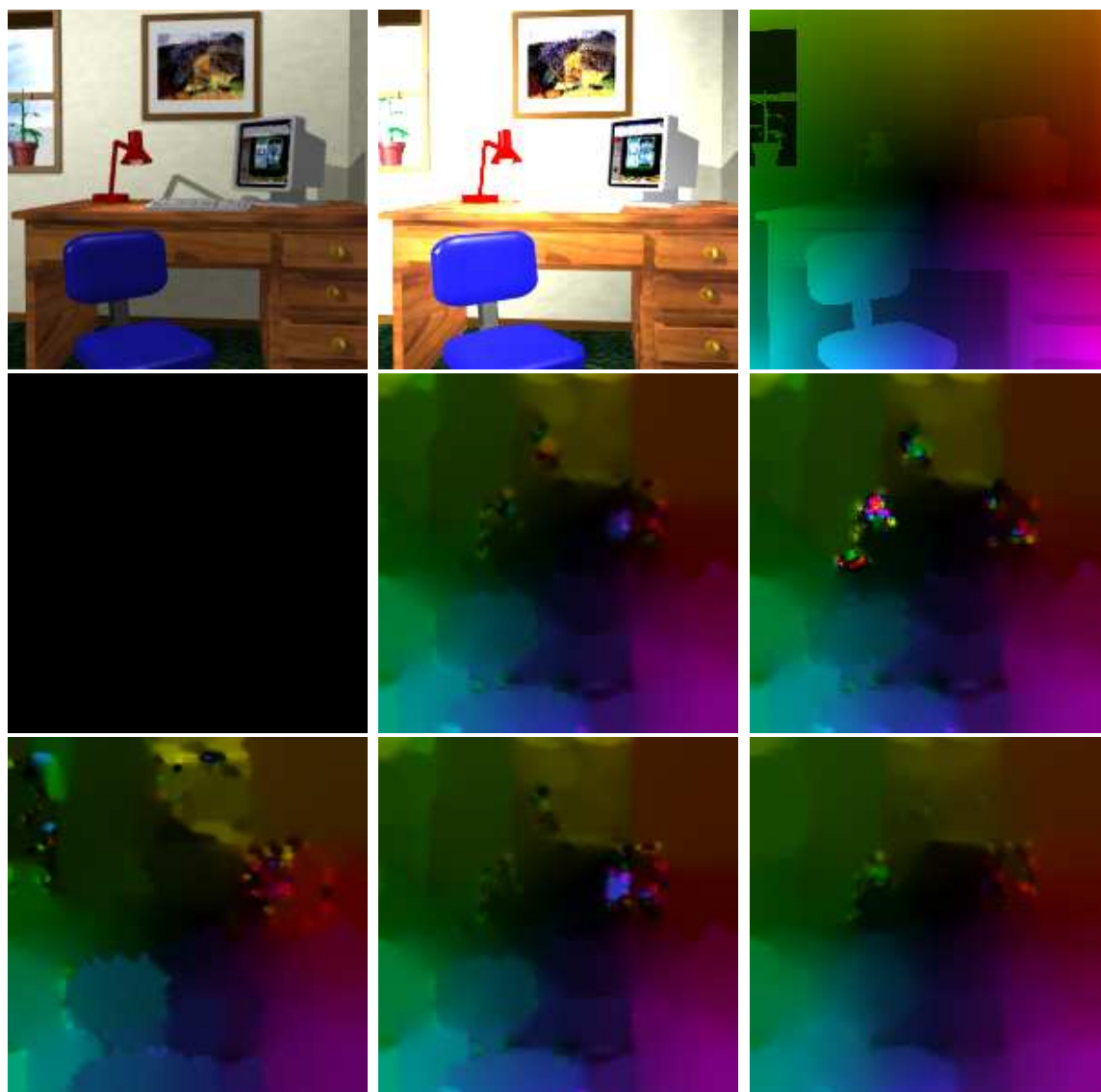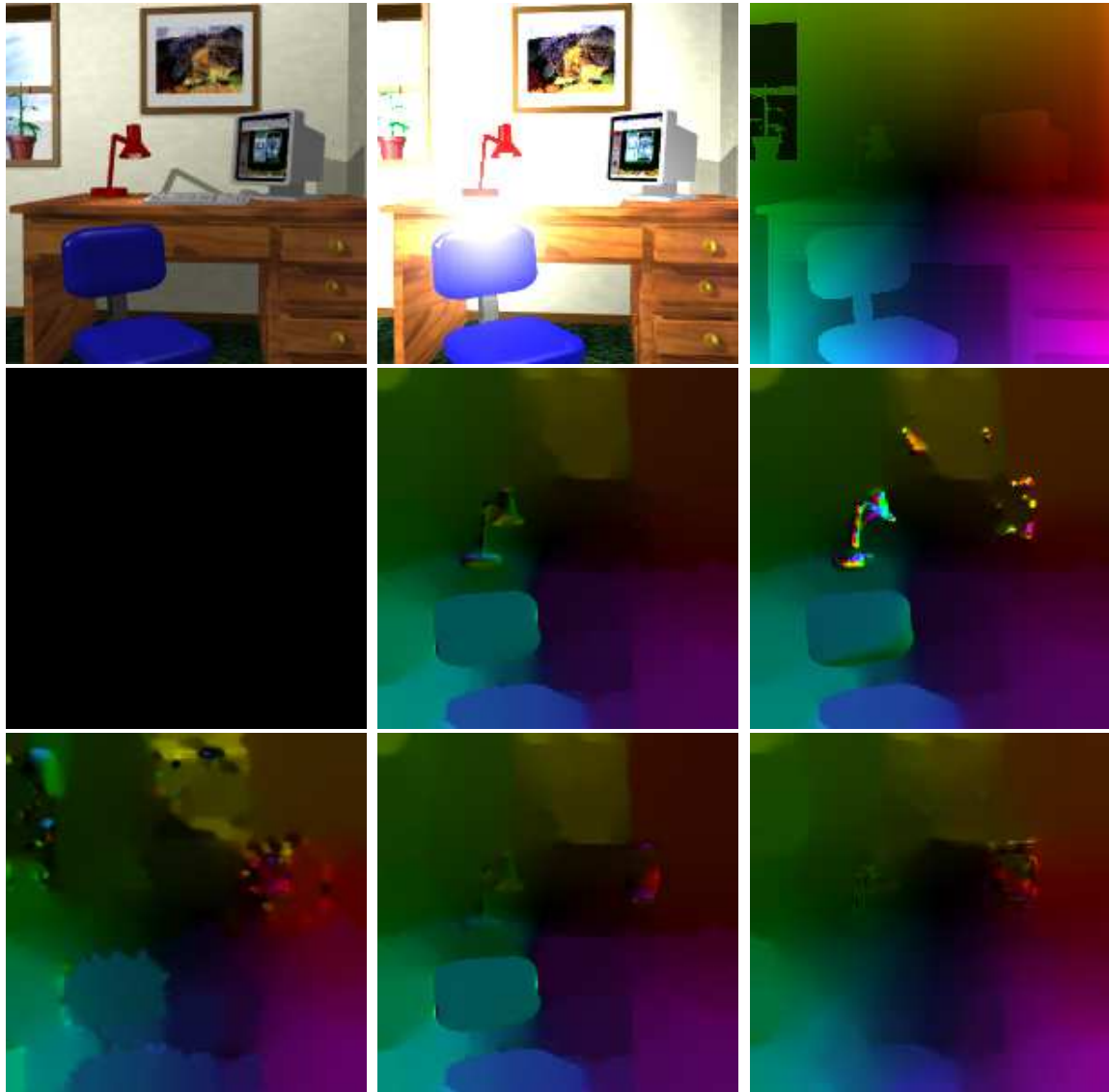
Figure 6.21: **Experiment 4.III:** Multiplicative and additive changes **Top Left:** Frame 10 of the "Office" sequence; **Top Center:** Frame 11 (with multiplicative and additive illumination changes); **Top Right:** The ground truth colour-coded flow field between frame 10 and frame 11; **Middle Left:** The colour-coded flow field, computed with the Brox *et al.* multichannel method for the RGB constancy assumption; **Middle Center:** The arithmetic mean assumption; **Middle Right:** The geometric mean assumption; **Bottom Left:** The hue assumption; **Bottom Center:** The $\phi\theta$ assumption; **Bottom Right:** The log-derivatives assumption.

6.2.3   Real-world sequence experiment

At the end of this section we will demonstrate the power of the photometric invariance assumptions, when used for a real-world image sequence with illumination changes introduced. As this is a real-world sequence, we do not have a ground truth available. In this case we can count only on the flow field representation, to compare our results.

For this experiment we will use only the Brox *et al.* multichannel method, as it is the better performing one. The computed flow fields can be seen on Figure 6.22.

Like we have discussed earlier, the motion in this sequence is divergent and based on this knowledge we can predict how the ideal optic flow would look like. The flow field that is given by the *RGB assumption* obviously is not a correct one, as it states that the road is moving in upward direction (the yellow colour means that the vectors point up). In comparison, our photometric invariants perform much better and produce a realistic flow field. Because, the displacements in this sequence are quite large (which can be seen from the two used frames), we have some problems to detect the small pedestrian in the sequence. The pedestrian is moving with a speed greater than the speed of the environment and that is why he should be detected and coloured properly. With our photometric assumptions (the *geometric mean* and the $\phi\theta$) we can detect him and accurately estimate his speed.

With this example we demonstrated that, when combining suitable photometric constancy assumptions with a good multichannel variational method, we can compute a highly robust optic flow with respect to severe realistic illumination changes.

6.3   Rotation Experiments

In the previous section of this chapter, we presented experiments that illustrate the usefulness of the photometric invariance assumptions. In this section we will focus our attention on the rotational moment invariants (see Section 4.6). These moments were derived in such a way as to compute accurate optic flow for sequences where we have rotational motion introduced.

We have used the colour rotational moments Φ_i (see Section 4.7) together with the multichannel Horn-Schunck and Brox *et al.* methods. Every moment has been computed in a 5×5 neighbourhood window. After examining the eight presented moments, we reached the conclusion that the best results are given by the 0th and the 1st moments - Φ_0 and Φ_1. These two moments discard least information and so it is not surprising that they perform better than the rest.

6.3.1   Slowly Rotating Disk experiment

Now, let us take a look at the first experiment. We chose the "Rotating Disk" sequence for the first rotation test. Our best results are listed in Table 6.5. All wights γ_i were
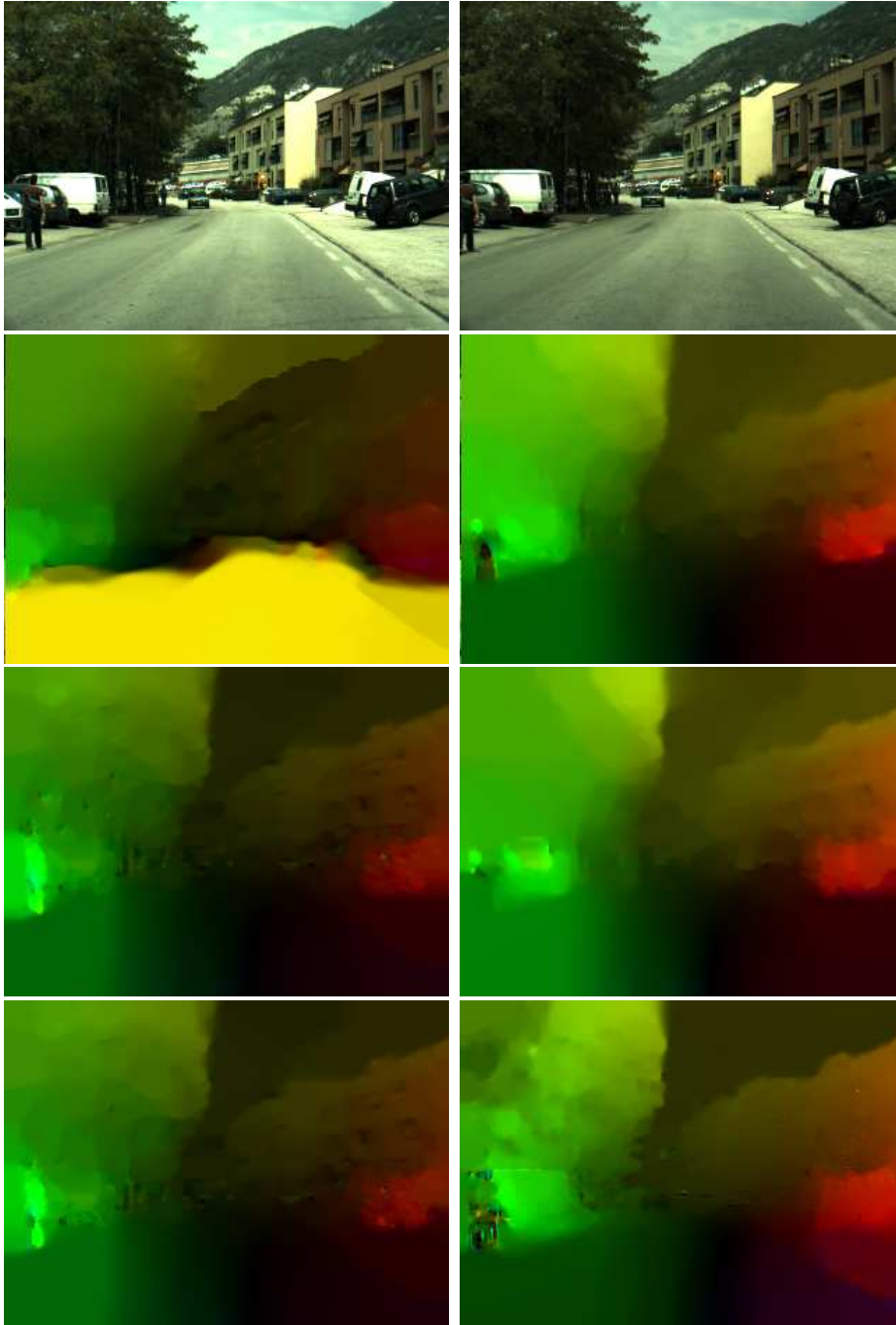
Figure 6.22: **Experiment 5:** Results under real illumination conditions **Top Left:** Left frame 205 of the Road sequence; **Top Right:** Left frame 207; **Upper Middle Left:** The RGB constancy assumption, Brox *et al.*; **Upper Middle Right:** The arithmetic mean assumption; **Lower Middle Left:** The geometric mean assumption; **Lower Middle Right:** The hue assumption; **Bottom Left:** The $\phi\theta$ assumption; **Bottom Right:** The log-derivatives assumption.

set to 1 and the rest of the parameters were optimized with respect to the AAE.

In the same table with the Φ_0 results, we have also provided the results for the *RGB constancy assumption*, when used for computing the optic flow for this sequence and the combined $RGB + \Phi_0$ results. As the reader has most probably noticed, the AAE of the *RGB* is better than the AAE of our best performing rotational moment. However, when we combine these two assumptions, we receive a much better result - AAE of 10.52^0 and 4.39^0 degrees. The best available result for this sequence, known to the author, is namely the *RGB* result [6]. This leads us to the conclusion that combining constancy assumptions in the data term, that contribute in a different, but still valid, manner, results in a very good practicle performance.

The actual computed flow fields are shown on Figure 6.23.

Concept	#Ch	Horn-Schunck multch.	Brox *et al.* multich.
RGB	3	10.63^0	4.68^0
Moment Φ_0	3	15.38^0	7.21^0
RGB + Φ_0	6	10.52^0	4.39^0

Table 6.5: *Performance of the 0th rotational moment Φ_0 for the "Disk" seq.*

6.3.2   Rotating Sphere experiment

With this second rotational experiment we would like to confirm our observations from the previous test. For our second experiment we chose the "Rotating Sphere" colour sequence.

The best results that we received are listed in Table 6.6. All wights γ_i were again set to 1 and the rest of the parameters were optimized with respect to the AAE. The results show the same tendency that we observed in Table 6.5. Again, the *RGB assumption*, with its AAE of 3.99^0 and 2.88^0 degrees, performs better than the *rotational moment assumption*, with its AAE of 6.65^0 and 4.29^0. However, when we combine these two assumptions we receive a much better result - AAE of 2.42^0 and 2.29^0 degrees. This once more proves that combining the rotational moments with other successful assumptions results in better performance with respect to the average angular error.

Our considerations are also confirmed by the corresponding flow fields shown on Figure 6.24.

Concept	#Ch	Horn-Schunck multch.	Brox *et al.* multich.
RGB	3	3.99^0	2.88^0
Moment Φ_0	3	6.65^0	4.29^0
RGB + Φ_0	6	2.42^0	2.29^0

Table 6.6: *Performance of the 0th rotational moment Φ_0 for the "Sphere" seq.*
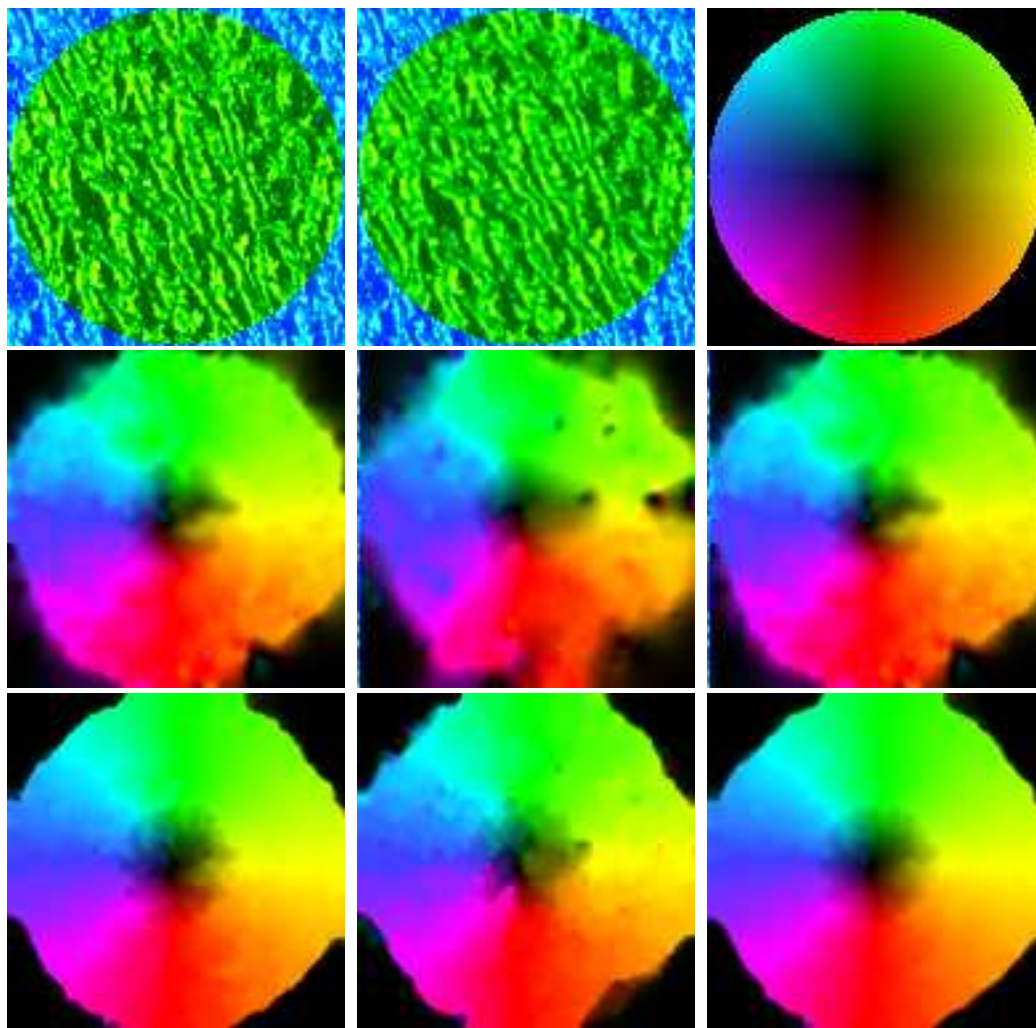
Figure 6.23: **Experiment 7:** Results under rotational motion for the "Slowly Rotating Disk" sequence **Top Left:** Frame 1 of the "Disk" sequence; **Top Center:** Frame 2 of the "Disk" sequence; **Top Right:** Colour plot of the ground truth between the 1st and the 2nd frames of the sequence; **Middle Left:** The RGB assumption, computed with the Horn-Schunck multichannel method; **Middle Center:** Moment Φ_0, Horn-Schunck; **Middle Right:** $RGB + \Phi_0$, Horn-Schunck; **Bottom Left:** The RGB assumption, computed with the Brox *et al.* multichannel method; **Bottom Center:** Moment Φ_0, Brox *et al.* **Bottom Right:** $RGB + \Phi_0$, Brox *et al.*
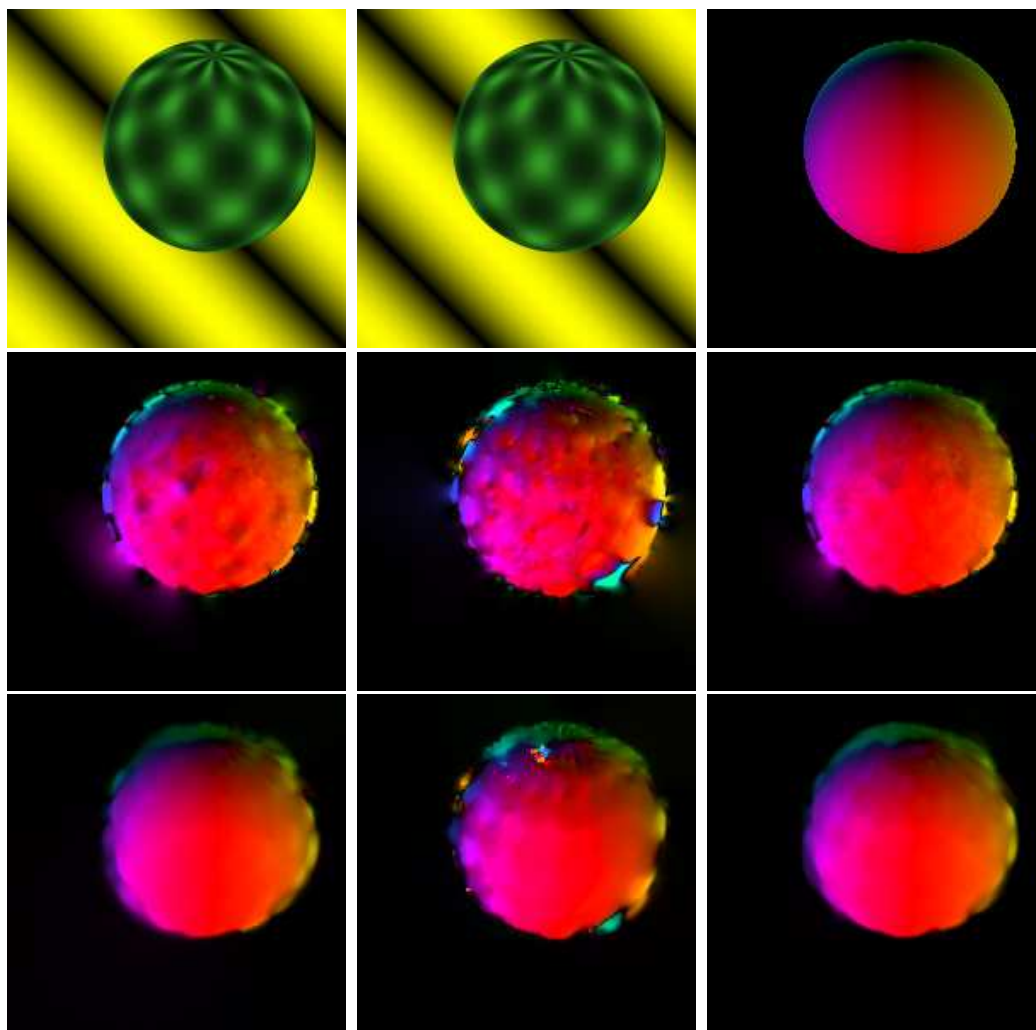
Figure 6.24: **Experiment 6:** Results under rotational motion for the "Rotating Sphere" sequence **Top Left:** Frame 10 of the "Sphere" sequence; **Top Center:** Frame 11; **Top Right:** Colour plot of the ground truth between the 10th and the 11th frames of the sequence; **Middle Left:** The RGB assumption, computed with the Horn-Schunck multichannel method; **Middle Center:** Moment Φ_0, Horn-Schunck; **Middle Right:** $RGB + \Phi_0$, Horn-Schunck; **Bottom Left:** The RGB assumption, computed with the Brox *et al.* multichannel method; **Bottom Center:** Moment Φ_0, Brox *et al.* **Bottom Right:** $RGB + \Phi_0$, Brox *et al.*

6.4 Summary

In this last chapter we have evaluated the performance of the photometric, as well as, of the rotational moment invariants using famous colour image test sequences.

Based on the estimated results, we would say that when having realistic illumination changes introduced, the combination of photometric invariance assumptions with multichannel variational methods provides better performance and gives stable and reliable results.

On the other hand, when rotational motion is introduced, it pays off to combine rotationally invariant assumptions, like the rotational moments from Chapter 4, with other well-performing assumptions in order to receive better results as an outcome.

Chapter 7

Summary and Conclusions

The variational methods are the most successful methods used for optic flow computation. One of the contributions of this work is the extending of the famous methods of Horn-Schucnk and Brox *et al.*, to methods that allows us to combine multiple constancy assumptions in the data term. We have extended these methods in Chapter 5 in order to be able to combine invariant information from multiple channels and thus to compute a better optic flow solution.

The main goal of all the research done in optic flow computation is the finding of a correct optic flow for real-world data. One of the main challenges when it comes to real-world images is the presence of illumination changes. These illumination changes may spoil the final result if they are not considered and not processed in a correct manner. What we have tried to do in Chapter 3 of this work is to find suitable assumptions that can be used in the energy functional's data term, in order to combat these illumination changes. In the Experimental part we have showed that these assumptions work for synthetic as well as for real-world data.

Another challenge that stays in front of the optic flow methods is the correct treatment of motion patterns. The second type of constancy assumptions that we have explored deals with the detection of rotational motion. In Chapter 4 we have tried to modify the statistical moments in such a way as to make them suitable for usage for optic flow computation and invariant under rotational motion. As we have demonstrated in the Experiments section of this work, the moments can be quite useful and can contribute a lot for the computation of a better and more correct flow field for sequences with rotational motion introduced.

Both of the above mentioned challenges, as well as their effect on the finding of a better optic flow solution, have not been explored in detail so far in the literature. We hope that with our work we have managed to connect the worlds of photometric and rotational invariance with the world of variational methods. We further hope that our efforts have contributed for the improvement of the highly accurate variational methods and the optic flow computation techniques.

Appendix A

In Chapter 4 we have presented rotation invariant moments (4.7), which are not based on the definition of the central moments. Here we would like to prove that these moments are truly rotationally invariant.

In Chapter 4 we derived these rotation invariant moments for a discrete image, based on the generalized colour moments (4.5) and the scaled moments (4.6). To make our proof more general we will perform it for continuous images and, for the purpose of simplicity, for black-and-white images.

This proof can be applied straightforwardly for the discrete colour images afterwards.

For a continuous image $f : \mathbb{R}^2 \to \mathbb{R}$ one defines the moments of order $p + q$ as:

$$m_{p,q} = \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} x^p \, y^q \, f(x, y) \, dx \, dy \tag{A.1}$$

for $p, q = 0, 1, 2, \ldots$

We can skip the scaling step, as it doesn't change anything in the proof.

So, we have the following rotation invariant moments:

$$
\begin{aligned}
\Phi_0 &= m_{1,0}^2 + m_{0,1}^2 \tag{A.2}\\
\Phi_1 &= m_{2,0} + m_{0,2}\\
\Phi_2 &= (m_{2,0} - m_{0,2})^2 + 4m_{1,1}^2\\
\Phi_3 &= (m_{3,0} - 3m_{1,2})^2 + (3m_{2,1} - m_{0,3})^2\\
\Phi_4 &= (m_{3,0} + m_{1,2})^2 + (m_{2,1} + m_{0.3})^2\\
\Phi_5 &= (m_{3,0} - 3m_{1,2})(m_{3,0} + m_{1,2})[(m_{0,3} + m_{1,2})^2 - 3(m_{2,1} + m_{0,3})^2] +\\
&\quad (3m_{2,1} - m_{0,3})(m_{2,1} + m_{0,3})[3(m_{3,0} + m_{1,2})^2 - (m_{2,1} + m_{0,3})^2]\\
\Phi_6 &= (m_{2,0} - m_{0,2})[(m_{3,0} + m_{1,2})^2 - (m_{2,1} + m_{0,3})^2] +\\
&\quad 4m_{1,1}(m_{3,0} + m_{1,2})(m_{2,1} + m_{0,3})\\
\Phi_7 &= (3m_{1,2} - m_{3,0})(m_{3,0} + m_{1,2})[(m_{3,0} + m_{1,2})^2 - 3(m_{2,1} + m_{0,3})^2] +\\
&\quad (3m_{2,1} - m_{0,3})(m_{2,1} + m_{0,3})[3(m_{3,0} + m_{1,2})^2 - (m_{2,1} + m_{0,3})^2]
\end{aligned}
$$

The following proofs are based on the theory of polar coordinates, where we have that:

$$
\begin{aligned}
x &= r\cos\theta \\
y &= r\sin\theta \\
r &= \sqrt{x^2 + y^2}
\end{aligned}
\tag{A.3}
$$

Please note that r and θ, as defined, depend on x and y.

Proof of Φ_0

$$
\begin{aligned}
\Phi_0 &= m_{1,0}^2 + m_{0,1}^2 \\
&= \left(\iint x\, f(x,y)\, dx\, dy \right)^2 + \left(\iint y\, f(x,y)\, dx\, dy \right)^2 \\
&= \iiiint x\, \bar{x}\, f\, \bar{f}\, dx\, dy\, d\bar{x}\, d\bar{y} + \iiiint y\, \bar{y}\, f\, \bar{f}\, dx\, dy\, d\bar{x}\, d\bar{y} \\
&= \iiiint (x\bar{x} + y\bar{y})\, \bar{f}f\, dx\, dy\, d\bar{x}\, d\bar{y} \\
&= \iiiint (r\,\cos\theta_1\, \bar{r}\cos\theta_2 + r\,\sin\theta_1\, \bar{r}\sin\theta_2)\, \bar{f}f\, dx\, dy\, d\bar{x}\, d\bar{y} \\
&= \iiiint [\bar{r}\, r\, (\cos\theta_1\, \cos\theta_2 + \sin\theta_1\, \sin\theta_2)]\, \bar{f}\, f\, dx\, dy\, d\bar{x}\, d\bar{y} \\
&= \iiiint \bar{r}\, r\, \cos(\theta_1 - \theta_2)\, \bar{f}\, f\, dx\, dy\, d\bar{x}\, d\bar{y}
\end{aligned}
\tag{A.4}
$$

where $f\bar{f}r\bar{r}\cos(\theta_1 - \theta_2)$ is rotationally invariant, as the difference between the angles θ_1 and θ_2 does not change under rotation. This makes Φ_0 also rotationally invariant.

Proof of Φ_1

$$
\begin{aligned}
\Phi_1 &= m_{2,0} + m_{0,2} \\
&= \iint x^2\, f\, dx\, dy + \iint y^2\, f\, dx\, dy \\
&= \iint (x^2 + y^2)\, f\, dx\, dy \\
&= \iint r^2\, f\, dx\, dy
\end{aligned}
\tag{A.5}
$$

which is rotationally invariant

Proof of Φ_2

$$
\begin{aligned}
\Phi_2 &= (m_{2,0} - m_{0,2})^2 + 4m_{1,1}^2 \tag{A.6}\\
&= \left(\iint x^2 \, f \, dx \, dy - \iint y^2 \, f \, dx \, dy \right)^2 + 4\left(\iint xy \, f \, dx \, dy \right)^2 \\
&= \left(\iint (x^2 - y^2) \, f \, dx \, dy \right)^2 + 4\left(\iint xy \, f \, dx \, dy \right)^2 \\
&= \iiiint (x^2 - y^2)(\bar{x}^2 - \bar{y}^2) \, f\bar{f} \, dx \, dy \, d\bar{x} \, d\bar{y} + 4\iiiint xy \, \bar{x}\bar{y} \, f\bar{f} \, dx \, dy \, d\bar{x} \, d\bar{y} \\
&= \iiiint f\bar{f}\left((x^2 - y^2)(\bar{x}^2 - \bar{y}^2) + 4xy\bar{x}\bar{y} \right) dx \, dy \, d\bar{x} \, d\bar{y} \\
&= \iiiint f\bar{f}\left((x\bar{x} + y\bar{y})^2 - (x\bar{y} - \bar{x}y)^2 \right) dx \, dy \, d\bar{x} \, d\bar{y} \\
&= \iiiint f\bar{f}\left((r\bar{r}cos(\theta_1 - \theta_2))^2 - (r\bar{r}sin(\theta_1 - \theta_2))^2 \right) dx \, dy \, d\bar{x} \, d\bar{y} \\
&= \iiiint f\bar{f}r^2\bar{r}^2(cos^2(\theta_1 - \theta_2) - sin^2(\theta_1 - \theta_2)) \, dx \, dy \, d\bar{x} \, d\bar{y} \\
&= \iiiint f\bar{f}r^2\bar{r}^2 cos2(\theta_1 - \theta_2) \, dx \, dy \, d\bar{x} \, d\bar{y}
\end{aligned}
$$

where $f\bar{f}r^2\bar{r}^2 cos\,2(\theta_1 - \theta_2)$ is rotationally invariant, which makes Φ_2 also rotationally invariant.

Proof of Φ_3

$$
\begin{aligned}
\Phi_3 &= (m_{3,0} - 3m_{1,2})^2 + (3m_{2,1} - m_{0,2})^2 \tag{A.7}\\
&= \left(\iint x^3 f \, dx \, dy - 3\iint xy^2 f \, dx \, dy \right)^2 + \left(3\iint x^2 y f \, dx \, dy - \iint y^3 f \, dx \, dy \right)^2 \\
&= \left(\iint (x^3 - 3xy^2) f \, dx \, dy \right)^2 + \left(\iint (3x^2 y - y^3) f \, dx \, dy \right)^2 \\
&= \iiiint (x^3 - 3xy^2)(\bar{x}^3 - 3\bar{x}\bar{y}^2) f\bar{f} \, dx \, dy \, d\bar{x} \, d\bar{y} + \\
&\quad + \iiiint (3x^2 y - y^3)(3\bar{x}^2\bar{y} - \bar{y}^3) f\bar{f} \, dx \, dy \, d\bar{x} \, d\bar{y} \\
&= \iiiint f\bar{f}\Big(x^3\bar{x}^3 - 3x^3\bar{x}\bar{y}^2 - 3x\bar{x}^3 y^2 + 9x\bar{x}y^2\bar{y}^2 + \\
&\quad + 9x^2\bar{x}^2 y\bar{y} - 3x^2 y\bar{y}^3 - 3\bar{x}^2 y^3\bar{y} + y^3\bar{y}^3 \Big) dx \, dy \, d\bar{x} \, d\bar{y}
\end{aligned}
$$

$$= \iiiint f\bar{f}\Big((x\bar{x} + y\bar{y})^3 - 3x\bar{x}(x\bar{y} - \bar{x}y)^2 - 3y\bar{y}(x\bar{y} - \bar{x}y)^2\Big)\, dx\, dy\, d\bar{x}\, d\bar{y}$$

$$= \iiiint f\bar{f}\Big((x\bar{x} + y\bar{y})((x\bar{x} + y\bar{y})^2 - 3(x\bar{y} - \bar{x}y)^2)\Big)\, dx\, dy\, d\bar{x}\, d\bar{y}$$

$$= \iiiint f\bar{f}r^3\bar{r}^3 cos(\theta_1 - \theta_2)\Big(cos^2(\theta_1 - \theta_2) - 3sin^2(\theta_1 - \theta_2)\Big)\, dx\, dy\, d\bar{x}\, d\bar{y}$$

$$= \iiiint f\bar{f}r^3\bar{r}^3 cos(\theta_1 - \theta_2)(2\,cos\,2(\theta_1 - \theta_2) - 1)\, dx\, dy\, d\bar{x}\, d\bar{y}$$

where $f\bar{f}r^3\bar{r}^3 cos(\theta_1 - \theta_2)(2\,cos\,2(\theta_1 - \theta_2) - 1)$ is rotationally invariant, which makes Φ_3 also rotationally invariant.

Proof of Φ_4

$$\begin{aligned}
\Phi_4 &= (m_{3,0} + m_{1,2})^2 + (m_{2,1} + m_{0,3})^2 \qquad\qquad\qquad\qquad\qquad \text{(A.8)}\\
&= \left(\iint (x^3 + xy^2)f\, dx\, dy\right)^2 + \left(\iint (x^2y + y^3)f\, dx\, dy\right)^2\\
&= \iiiint f\bar{f}(x^3 + xy^2)(\bar{x}^3 + \bar{x}\bar{y}^2)\, dx\, dy\, d\bar{x}\, d\bar{y}\,+\\
&\quad + \iiiint f\bar{f}(x^2y + y^3)(\bar{x}^2\bar{y} + \bar{y}^3)\, dx\, dy\, d\bar{x}\, d\bar{y}\\
&= \iiiint f\bar{f}\Big(x^3\bar{x}^3 + x^3\bar{x}\bar{y}^2 + x\bar{x}^3y^2 + x\bar{x}y^2\bar{y}^2 + x^2\bar{x}^2y\bar{y} + x^2y\bar{y}^3 + \bar{x}^2\bar{y}y^3 + y^3\bar{y}^3\Big)\\
&= \iiiint f\bar{f}\Big((x\bar{x} + y\bar{y})^3 + x\bar{x}(x\bar{y} - \bar{x}y)^2 + y\bar{y}(x\bar{y} - \bar{x}y)^2\Big)\, dx\, dy\, d\bar{x}\, d\bar{y}\\
&= \iiiint f\bar{f}(x\bar{x} + y\bar{y})\Big((x\bar{x} + y\bar{y})^2 + (x\bar{y} - \bar{x}y)^2\Big)\, dx\, dy\, d\bar{x}\, d\bar{y}\\
&= \iiiint f\bar{f}r\bar{r}cos(\theta_1 - \theta_2)\Big(r^2\bar{r}^2 cos^2(\theta_1 - \theta_2) + r^2\bar{r}^2 sin^2(\theta_1 - \theta_2)\Big)\, dx\, dy\, d\bar{x}\, d\bar{y}\\
&= \iiiint f\bar{f}r^3\bar{r}^3 cos(\theta_1 - \theta_2)\, dx\, dy\, d\bar{x}\, d\bar{y}
\end{aligned}$$

where $f\bar{f}r^3\bar{r}^3 cos\,(\theta_1 - \theta_2)$ is rotationally invariant, which makes Φ_4 also rotationally invariant.

Proofs of Φ_5, Φ_6 and Φ_7   The proofs of the last three rotationally invariant moments are analogous to the previous proofs.

Bibliography

[1] J. L. Barron, D. J. Fleet, and S. S. Beauchemin. Performance of optical flow techniques. *International Journal of Computer Vision*, 12(1):43–77, February 1994.

[2] M. Bertero, T. A. Poggio, and V. Torre. Ill-posed problems in early vision. *Proceedings of the IEEE*, 76(8):869–889, August 1988.

[3] M. J. Black and P. Anandan. The robust estimation of multiple motions: parametric and piecewise smooth flow fields. *Computer Vision and Image Understanding*, 63(1):75–104, January 1996.

[4] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High accuracy optic flow estimation based on a theory for warping. In T. Pajdla and J. Matas, editors, *Computer Vision – ECCV 2004*, volume 3024 of *Lecture Notes in Computer Science*, pages 25–36. Springer, Berlin, 2004.

[5] T. Brox, A. Bruhn, and J. Weickert. Variational motion segmentation with level sets. In H. Bischof, A. Leonardis, and A. Pinz, editors, *Computer Vision – ECCV 2006*, volume 3951 of *Lecture Notes in Computer Science*, pages 471–483. Springer, Berlin, 2006.

[6] A. Bruhn. *Variational Optic Flow Computation – Accurate Modelling and Efficient Numerics*. PhD thesis, Department of Mathematics and Comptuer Science, Saarland University, Germany, July 2006.

[7] A. Bruhn and J. Weickert. Towards ultimate motion estimation: Combining highest accuracy with real-time performance. In *Proc. Tenth International Conference on Computer Vision*, volume 1, pages 749–755, Beijing, China, June 2005. IEEE Computer Society Press.

[8] A. Bruhn, J. Weickert, C. Feddern, T. Kohlberger, and C. Schnörr. Variational optical flow computation in real-time. *IEEE Transactions on Image Processing*, 14(5):608–615, May 2005.

[9] A. Bruhn, J. Weickert, T. Kohlberger, and C. Schnörr. A multigrid platform for real-time motion computation with discontinuity-preserving variational methods. *International Journal of Computer Vision*, 70(3):257–277, December 2006.

[10] G. D. Finlayson, M. Drew, and C. Lu. Intrinsic images by entropy minimization. In T. Pajdla and J. Matas, editors, *Computer Vision – ECCV 2004*, volume 3023 of *Lecture Notes in Computer Science*, pages 582–595. Springer, Berlin, 2004.

[11] G. D. Finlayson and M. S. Drew. 4-Sensor camera calibration for image representation invariant to shading, shadows, lighting, and specularities. In *ICCV'01: International Conference on Computer Vision*, volume II, pages 473–480, 2001.

[12] I. M. Gelfand and S. V. Fomin. *Calculus of Variations*. Dover, New York, 2000.

[13] P. Golland and A. M. Bruckstein. Motion from color. *Computer Vision and Image Understanding*, 68(3):346–362, December 1997.

[14] R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Addison–Wesley, Reading, second edition, 2002.

[15] W. Hackbusch. *Iterative Solution of Large Sparse Systems of Equations*, volume 95 of *Applied Mathematical Sciences*. Springer, New York, 1993.

[16] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. Springer, 2001.

[17] B. Horn and B. Schunck. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.

[18] M-K. Hu. Visual pattern recognition by moment invariants. *IRE Trans. on Information Theory*, IT-8:pp:179–187, 1962.

[19] Y.-H. Kim, A. M. Martínez, and A. C. Kak. Robust motion estimation under varying illumination. *Image and Vision Computing*, 23(1):365–375, 2005.

[20] G. J. Klinker, S. A. Shafer, and T. Kanade. Color image analysis with an intrinsic reflection model, 1988.

[21] G. J. Klinker, S. A. Shafer, and T. Kanade. The measurement of highlights in color images. *International Journal of Computer Vision*, 2:7–32, 1988.

[22] G. J. Klinker, S. A. Shafer, and T. Kanade. A physical approach to color image understanding. *IJCV*, 4(1. January 1990):7–38, January 1990.

[23] H. C. Lee, E. J. Breneman, and C. P. Schulte. Modeling light reflection for computer vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12:402–409, 1990.

[24] B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proc. Seventh International Joint Conference on Artificial Intelligence*, pages 674–679, Vancouver, Canada, August 1981.

[25] *MathWorld.* `http://mathworld.wolfram.com/`, Last accessed: May 2007.

[26] B. McCane, K. Novins, D. Crannitch, and B. Galvin. On benchmarking optical flow. *Comput. Vis. Image Underst.*, 84(1):126–143, 2001.

[27] E. Mémin and P. Pérez. Dense estimation and object-based segmentation of the optical flow with robust techniques. *IEEE Transactions on Image Processing*, 7(5):703–719, May 1998.

[28] E. Mémin and P. Pérez. Hierarchical estimation and segmentation of dense motion fields. *International Journal of Computer Vision*, 46(2):129–155, 2002.

[29] F. Mindru, T. Tuytelaars, L.V. Gool, and Th. Moons. Moment invariants for recognition under changing viewpoint and illumination. In *Computer Vision and Image Understanding*, volume 94, pages 3–27, 2004.

[30] H.-H. Nagel and W. Enkelmann. An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8:565–593, 1986.

[31] S. Negahdaripour. Revised definition of optical flow: integration of radiometric and geometric clues for dynamic scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(9):961–979, 1998.

[32] N. Ohta. Optical flow detection by color images. In *Proc. Tenth International Conference on Pattern Recognition*, pages 801–805, Singapore, September 1989.

[33] N. Papenberg, A. Bruhn, T. Brox, S. Didas, and J. Weickert. Highly accurate optic flow computation with theoretically justified warping. *International Journal of Computer Vision*, 67(2):141–158, April 2006.

[34] S. A. Shafer. Using color to seperate reflection components. *Color Research and Applications*, 10(4):210–218, 1985.

[35] R. T. Tan, K. Nishino, and K. Ikeuchi. Illumination chromaticity estimation using inverse-intensity chromaticity space. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 673–680, 2003.

[36] Shoji Tominaga. Surface identification using the dichromatic reflection model. *IEEE Trans. Pattern Anal. Mach. Intell.*, 13(7):658–670, 1991.

[37] S. Uras, F. Girosi, A. Verri, and V. Torre. A computational approach to motion perception. *Biological Cybernetics*, 60:79–87, 1988.

[38] J. van de Weijer and Th. Gevers. Robust optical flow from photometric invariants. In *Proc. 2004 IEEE International Conference on Image Processing*, volume 3, pages 1835–1838, Singapore, October 2004.

[39] J. Weickert, A. Bruhn, and C. Schnörr. Lucas/Kanade meets Horn/Schunck: Combining local and global optic flow methods. Technical Report 17/2001, Computer Science Series, University of Mannheim, Germany, September 2001.

[40] J. Weickert and C. Schnörr. Variational optic flow computation with a spatio-temporal smoothness constraint. *Journal of Mathematical Imaging and Vision*, 14(3):245–255, May 2001.

[41] *Wikipedia, the free encyclopedia.* `http://www.wikipedia.org/`, Last accessed: May 2007.