Occupant Detection using Computer Vision

Marcus Klomark

LiTH-ISY-EX-3026

2000-05-19
Occupant Detection using Computer Vision

Master’s Thesis in Computer Vision at Linköping University

by

Marcus Klomark

LiTH-ISY-EX-3026

Examiner: Klas Nordberg

Linköping, May 2000
Abstract

The purpose of this master’s thesis was to study the possibility to use computer vision methods to detect and classify objects in the front passenger seat in a car. This work presents different approaches to solve this problem and evaluates the usefulness of each technique. The classification information should later be used to modulate the speed and the force of the airbag, to be able to provide each occupant with optimal protection and safety.

This work shows that computer vision has a great potential in order to provide data, which may be used to perform reliable occupant classification. Future choice of method to use depends on many factors, for example costs and requirements on the system from laws and car manufacturers. Further, evaluation and tests of the methods in this thesis, other methods, the ABE approach and post-processing of the results should also be made before a reliable classification algorithm may be written.

Keywords

Occupant detection, passenger classification, adaptive airbags, stereo vision, motion, colour vision, face detection.
Acknowledgements

I would like to thank Salah Hadi, Stina Hildemar, Björn Johansson, Jan-Olof Johansson, Karl Munsin, Klas Nordberg, Bengt-Arne Nyman, Anders Olsson and Håkan Prytz for useful help and interesting discussions.
## Contents

1 INTRODUCTION........................................................................................................1

1.1 AUTOLIV’S FRONTAL AIRBAG SYSTEM ..........................................................2

1.2 STATISTICS....................................................................................................3

1.3 PREVIOUS OCCUPANT PROTECTION EFFORTS ........................................4

1.3.1 Autoliv’s Adaptive Airbag system ........................................................4

1.3.2 Previous study in optical passenger detection .....................................5

1.4 PROBLEM DESCRIPTION ...........................................................................6

1.5 PROBLEM CONDITIONS ............................................................................7

1.5.1 Occupant variation problems...............................................................7

1.5.2 Lightning variation problems ..............................................................8

1.5.3 Advantages of the car environment......................................................8

1.5.4 Assumptions and delimitation ..............................................................8

1.5.5 Equipment ............................................................................................9

1.6 REPORT STRUCTURE ............................................................................10

2 SIGNIFICANT FEATURES........................................................................11

3 THREE DIMENSIONAL VISION ..............................................................13

3.1 STEREO VISION ....................................................................................14

3.1.1 Disparity and depth ...........................................................................16

3.1.2 The correspondence problem ............................................................16

3.1.3 Feature-based stereo vision ...............................................................18

3.1.4 Area-based stereo vision ....................................................................18

3.2 USING STEREO VISION IN THIS APPLICATION .....................................19

3.2.1 Normalised averaging ........................................................................19

3.2.2 Choice of applicability function .........................................................19

3.2.3 Results ................................................................................................21

3.2.4 Post-processing of the disparity map................................................22
3.2.5 Conclusions stereo vision ................................................................. 23
3.3 STRUCTURED LIGHT ........................................................................ 24
  3.3.1 Conclusions structured light .......................................................... 24
3.4 CONCLUSIONS USING 3D FEATURES .............................................. 26
  3.4.1 Classes that possibly can be detected .......................................... 26

4 MOTION .................................................................................................. 27
  4.1 OPTICAL FLOW .................................................................................. 28
    4.1.1 The aperture problem ................................................................. 29
    4.1.2 Optical flow using second derivatives ....................................... 30
  4.2 USING OPTICAL FLOW IN THIS APPLICATION ................................. 31
    4.2.1 Conclusions optical flow .............................................................. 32
  4.3 CONCLUSIONS USING MOTION ...................................................... 33
    4.3.1 Classes that possibly can be detected .......................................... 33

5 COLOUR VISION .................................................................................... 35
  5.1 COLOUR MODELS ............................................................................ 35
    5.1.1 The RGB colour model .............................................................. 36
    5.1.2 The HSV colour model .............................................................. 37
  5.2 SEGMENTATION BY SKIN COLOUR .................................................. 38
    5.2.1 Skin colour segmentation using HSV space ................................. 38
    5.2.2 Skin colour segmentation using other spaces ............................. 39
  5.3 USING HS SEGMENTATION IN THIS APPLICATION ......................... 40
    5.3.1 Hue segmentation ...................................................................... 40
    5.3.2 Saturation segmentation ............................................................ 42
    5.3.3 Hue and saturation segmentation .............................................. 42
    5.3.4 Post-processing of skin colour areas ......................................... 43
    5.3.5 Using HS segmentation as a data reduction ............................... 45
    5.3.6 Conclusions using HS segmentation ......................................... 46
  5.4 CONCLUSIONS USING COLOUR VISION ...................................... 48
    5.4.1 Classes that possibly can be detected .......................................... 48

6 FACE DETECTION ................................................................................. 49
  6.1 FINDING A FACE BY MOTION .......................................................... 50
1 Introduction

“...current air bags have been shown to be highly effective in reducing overall fatalities, they sometimes cause fatalities to out-of-position occupants, especially children. The agency's proposal would require that improvements be made in the ability of air bags to cushion and protect occupants of different sizes, belted and unbelted, and would require air bags to be redesigned to minimise risks to infants, children, and other occupants.”

NHTSA [1]

The text above is an extract from a new law proposal that has been set out by the American National Highway Traffic Safety Administration (NHTSA) [1], in order to reduce injuries caused by the expanding air bag and to enhance the benefits of air bags for all occupants. The proposal puts pressure on the air bag manufacturers to come up with solutions to the problem. The problem can be solved in arbitrary ways, which encourages innovative thinking.

Autoliv is a world-wide leader in automotive safety and a pioneer in air bags and Autoliv Design Centre in Linköping investigates whether a camera system in a car could be used to satisfy this law proposal. A classification of the occupant would be a good start to meet the demands of the law. So, if a reliable classification of the object in the passenger seat can be made, injuries caused by the expanding air bag can drastically be reduced.

---

1 An out-of-position occupant is inside the airbag expansion area.
1.1 Autoliv’s frontal airbag system

Autoliv’s frontal airbag system consists of sensors, an electronic control unit (ECU) and one airbag module or two, if the vehicle has a passenger bag. The airbag module, see Figure 1.1, consists of an inflator (or gas generator), an initiator and a cushion.

![Figure 1.1 A cross-section of a driver airbag module.](image1)

The sensors continuously monitor the acceleration and deceleration of the vehicle. This information is then sent to a microprocessor, where the crash algorithm of a vehicle is stored. When the crash pulse reaches the microprocessor an electrical current is sent to the airbag module’s initiator, which is a pyrotechnic unit that starts the inflation. The bag is fully inflated within 50 ms and deflated within 200 ms.

Even if the vehicle is equipped with an airbag, Autoliv points out that the seat belt is the primary mode of protection and always should be used. The seat belt stops the occupant from being propelled against the passenger airbag in a pre-crash braking and at the impact, see Figure 1.2.

![Figure 1.2 An unbelted occupant will be thrown forward in a frontal crash.](image2)
Introduction

1.2 Statistics

According to NHTSA [1]

- airbags have reduced fatalities in frontal crashes by about 30 percent and about 3,148 drivers and passengers have been saved, but
- 105 crashes in an airbag deployment had resulted in fatal injuries in United States since the introduction of airbags in 1986 until June 1, 1998.

Another conclusion is that these deaths did not occur randomly. The persons who have been killed or seriously injured by an airbag were extremely close to the airbag at the time of deployment. For some, this occurred because they were initially sitting too close to the airbag. For others because they were not restrained by seat belts or child safety seats and were thrown forward during pre-crash braking. The occupant will be hit by the airbag if it is too close to the cover of the airbag when the inflation starts, and the force can be so strong that the occupant will die.

There are three different types of occupants that are more likely to get killed or serious injured by the expanding airbag:

- unrestrained young children, who can easily be propelled close to or against the passenger airbag before the crash in a pre-crash braking
- infants in rear facing child seats, who ride with their heads extremely close to the passenger airbag
- drivers, especially unrestrained ones, who sit extremely close to the steering wheel (these drivers are most likely to be short women)

![Children travel safer in the rear seat than in the front seat.](image)

So, to travel safer in vehicles equipped with frontal airbags, the occupant should wear safety belt. Also, the driver should sit at least 10 inches away from the airbag, teenage and adult passengers can move their seat back and children should preferably ride in the rear seat, see Figure 1.3.
1.3 Previous occupant protection efforts

A temporary change to protect the occupants was to permit the vehicle manufacturers to install manual on-off switches for passenger airbags in vehicles without rear seats or with rear seats that are too small to accommodate a rear facing child restraint. Furthermore, improved labelling were required on new vehicles and child restraints, to inform and remind the driver and the occupants of the risks of putting a child seat in a front passenger seat, which is equipped with an airbag system. Other changes in progress are multiple level inflators and systems where the speed and the force of the airbag can be modulated, to provide each occupant with optimal protection and safety.

1.3.1 Autoliv's Adaptive Airbag system

Autoliv is developing a system that will vary the airbag deployment depending on the size, weight and dynamic position of an occupant and crash strength.

Figure 1.4 shows an overview of Autoliv’s Adaptive Airbag system with:
1. Ultra-sonic sensors
2. Weight sensors
3. Buckle sensors
4. Adaptive passenger inflator
5. Adaptive driver inflator
6. Seat position sensors
Introduction

This system has following functions:

- The ultra-sonic sensors gives information whether the occupant is out-of-position or not, and can be used to decide if the airbag should be deployed or not.
- The weight sensors give information about the weight of the occupant and the purpose of the weight system is to classify the occupant due to their weight, and can for example determine if the seat is empty or not.
- The buckle switch gives information whether the seat belt is used or not.
- The seat position sensors give information about how the seat is adjusted.

The airbag microprocessor will be able to regulate the inflation of the airbag to these different parameters. The Adaptive Airbags should fulfil a series of additional tests that are intended be phased in from model year 2003 in new light vehicles sold in the United States.

1.3.2 Previous study in optical passenger detection

Studies in the area of passenger detection using optical sensors have been done earlier by Autoliv SAGEM in France between 1995 and 1996 [26]. The conclusion at the end of this work was:

- Their detection principle, which was based on edges in images, was too poor.
- The camera, which was used, had to small image frame (64 x 64 pixels).
- The method required very good lightning.

One reason to look into this area again is that the cameras and processors constantly are improved, which enables usage of more powerful methods than they used.
1.4 Problem description

The objective of this thesis is to investigate the possibility to use computer vision methods on images, taken by cameras in a car, in order to classify objects in the passenger seat. These will be the main classes to distinguish between, see also Figure 1.5:
- Person
- Rear Faced Baby Seat (RFBS)
- Empty Seat

Certainly, the person class must be divided into different sub classes later like adults and children of different sizes, children in different front faced child restraints etc.

![Figure 1.5 Examples of the three occupant classes.](image)

The main part of the thesis gives reviews of different computer vision methods and discusses what good features these methods can extract, and not the classification part itself. This is meant to be a preliminary investigation and the purpose is to produce a summary of methods, which may produce proper results. Future choice of method/methods will be based on this thesis and therefore, it would be satisfying if as many appropriate methods as possible are investigated and compared to each other. Furthermore, if different independent methods are used for classification these can cooperate and complement each other and thus give an improved decision.

The main task is to find features that are significant for one object class and never, or hardly ever, appears in any other of the classes and to find methods to extract these features. Another important issue is to reduce the data, in a smart way to be able to speed up the process.
1.5 Problem conditions
Using a camera system in this application is possible because:
- The rapid increase in low cost computing power is now making the
development of real time computer vision systems practical.
- CMOS technique enables low cost cameras with high dynamic range,
random pixel access, on-chip signal processing etc.
- Camera manufacturers have already a great experience in mass-
producing good, cheap and small cameras, such as web-cameras.

So, these assumptions enables a solution to the problem, but lot of problems
have to be overcome to be able perform a robust classification.

1.5.1 Occupant variation problems
An intuitive problem is that occupants in the seat can differ a lot, even if
two occupants belong to the same class. Here are some examples:
- A person can be of different sizes, sexes, races etc. Numbers of sit poses
are huge. Some other variations for objects in this class are beard,
glasses, headgear and different haircuts and clothes.
- There is a wealth of different baby seats on the market that might be
faced rear, see Figure 1.6.
- There can be objects in the seat that not fits in any of the three classes,
for example a bag.

Also, the seat can be adjusted in many different positions. This introduces
further variations.

Figure 1.6 Some different child seats.
1.5.2 Lightning variation problems

Another huge challenge is that images of the scene will vary a lot because of shifting illumination. The different lightning conditions will provide great intensity variations in the images. In the dark, a lot of information is lost. For example, lines and edges are getting less sharp and the colour information in the images will disappear. If there is no light at all, the images will be totally black and no information can be extracted. Shades, bright spots cause specular reflection\(^1\) and strong illumination from the sun and headlights from other vehicles can also introduce problems.

1.5.3 Advantages of the car environment

So, there are a lot of problems to deal with, but the car environment also has its advantages. Some geometry of the scene is known. It might be a good thing to use prior knowledge about this.

1.5.4 Assumptions and delimitation

This work is an investigation and a whole functional system will not be developed. Not many qualitative results will not be presented, because there will not be enough time to implement the methods and make sufficiently large tests for this. Instead, possibilities will be pointed out often along with some illustrative examples and references to related works.

Time aspects is assumed to be of less importance in this work, because:

- The classification will not be very time dependant. The object in the passenger is not changing so fast during an ordinary trip, but a fast and robust system is of course better than a slow with the same robustness.
- A camera system for airbags will be in work earliest in 2005, so benefits can be obtained from Moore’s law (see Appendix).

---

\(^1\) Specular reflection appears at a certain viewing direction on an illuminated shiny surface, such as polished metal or a person’s face.
1.5.5 Equipment

Two different types of cameras generate test images, see Figure 1.7. Mitsubishi’s artificial retina sensor generated the image sequences.

- Resolution: 128 x 128 CMOS pixels
- Image area: 3.07 mm x 3.07 mm
- Pixel size: 24 $\mu$m x 24 $\mu$m

This camera is also sensitive in the IR spectrum. Other images were produced by one Olympus CAMEDIA C-830L digital camera, which has a maximum resolution of 1280 x 960 CCD pixels.

![Figure 1.7 Cameras used in this work.](image1)

Figure 1.7 Cameras used in this work.

A special camera stand was developed, see Figure 1.8, to be able to get good test images. Two cameras can be installed and they can be adjusted to point in arbitrary directions.

![Figure 1.8 A Mitsubishi camera attached to the camera stand.](image2)

Figure 1.8 A Mitsubishi camera attached to the camera stand.
1.6 Report structure

This report contains 9 chapters. This chapter was an introduction and gave the background to the problem. Chapter 2 discusses what significant features each occupant class has. The five proceeding chapters (3 to 7) will examine different areas of computer vision and give introductions to some different techniques in those areas. Suggestions of what features that might be used for occupant classification will be presented. Chapter 8 evaluates the different methods regarding usefulness, robustness etc. Chapter 9 gives a summary of the thesis.
2 Significant features

To pick methods for feature extraction we have to analyse what features that are likely to be useful. Some different features, which are significant to each class and assumed to be proper in differing objects in this class from other classes, will here be presented.

Person features:
• A person is always moving during a trip, at least sometimes.
• In most cases a camera will capture the face of a person.
• The 3D-shape of a person is ought to be a significant and useful feature.

Empty seat features:
• The 3D-shape is almost the same for all seats, even if the seat is various adjusted.
• No motion or face is present.
• The colour and texture of the seat is often continuous.

RFBS features:
• The 3D-shape differs a lot from the other classes.
• Colour, texture, lines and edges are some features that might be used to extract the form and 2D-shape of a RFBS.

Some methods that are able to produce these features will be presented in the proceeding chapters. Neural networks, which rely on other than by humans programmed algorithms, will also be discussed briefly.
3 Three dimensional vision

“The projection of light rays onto the retina presents our visual system with an image of the world that is inherently two-dimensional, yet we are able to interact with the three-dimensional world, even in situations new to us, or with objects unknown to us. That we accomplish this task easily implies that one of the functions of the human visual system is to reconstruct a 3-D representation of the world from its 2-D projection onto our eyes.”

Robyn Owens, Department of Computer Science, The University of Western Australia [2]

Most computer vision methods relies on two-dimensional information, say in x- and y-coordinates, and only works on the direct pixel intensity values. Such methods can easily get into problems, for example if a method have the purpose to detect faces of humans in a 3D scene, a picture on a wall containing a face can fool it. By introducing a further dimension, say depth in z-coordinates, this problem can be overcome.

Stereo vision will here be described more deeply than other methods. One reason why this method was chosen is that stereo is less dependent to changes in lightning than many other 3D methods, because the conditions are the same for the image pair, as the images are taken simultaneously. Further, the equipment for structured light was not available, but this technique will though be discussed briefly. Structure from motion will not be discussed, partly due to its sensitivity to changes in lightning and their complexity, but also because this method requires motion to produce 3D information.
3.1 Stereo vision

This method is very similar to human vision. We humans can use our two eyes to look at the world around us and our brain can combine the two slightly different views from each eye to produce three-dimensional perception. With the information from these three dimensions we are able to make judgements about distances, angles, shapes and volumes. By using a computer as a brain and two cameras, which are separated from each other, as eyes one can extract depth information in a scene.

Getting the third dimension is a matching problem. The aim is to find the same object or part of object in both images and then measuring the distance to it using triangulation. Matching objects at each pixel in the image leads to a so-called disparity map (see 2.1.2), from which a distance map can be generated.

We now look at a simplified stereo image system that is discussed by David Marshall [3].

![Figure 3.1 A simplified stereo imaging system.](image)

[Diagram of stereo imaging system showing object surface, baseline, and cameras at different distances]
Three dimensional vision

Figure 3.1 shows:
- Two cameras with their optical axes parallel and separated by a distance, \(d\).
- The line connecting the camera lens centres is called the baseline.
- Let the baseline be perpendicular to the line of sight of the cameras.
- Let the x-axis of the three-dimensional world coordinate system be parallel to baseline.
- Let the origin \(O\) of this system be mid-way between the lens centres.
- (x,y,z) is a point on an object surface.
- Let this point have image coordinates \((x_L,y_L)\) and \((x_R,y_R)\) in the left and right image planes of the perspective cameras.
- Let \(f\) be the focal length of both cameras. It is the perpendicular distance between the lens centre and the image plane.

Using similar triangles:

\[
\frac{x_L}{f} = \frac{x + d/2}{z} \quad \text{Equation 3-1}
\]

\[
\frac{x_R}{f} = \frac{x - d/2}{z} \quad \text{Equation 3-2}
\]

\[
\frac{y_L}{f} = \frac{y_R}{f} = \frac{y}{z} \quad \text{Equation 3-3}
\]

Solving for \((x,y,z)\) gives:

\[
x = \frac{d(x_L + x_R)}{2(x_L - x_R)} \quad \text{Equation 3-4}
\]

\[
y = \frac{d(y_L + y_R)}{2(y_L - y_R)} \quad \text{Equation 3-5}
\]

\[
z = \frac{df}{(x_L - x_R)} \quad \text{Equation 3-6}
\]
3.1.1 Disparity and depth

The disparity is:
- the relative displacement between two corresponding points
- equal to \((x_L - x_R)\) in Equation 3-4 and 3-6.
- inversely proportional to the depth (distance)
- proportional to the camera separation, \(d\)

The disparity map indicates where a pixel lies in the other image and, because of parallax, objects with different distances to the cameras will tend to have different displacements between the two images. An object positioned close to the camera is displaced more than an object far away, see Figure 3.2.

To get the depth, \(z\), of a pixel it must:
- be visible to both cameras
- be able to get identified in both pictures.

The accuracy of depth determination will increase with the camera separation, \(d\), but matching will be more difficult, because the differences in the both pictures will be larger.

3.1.2 The correspondence problem

The big challenge in stereo vision is to find corresponding regions in the images and, as shown, the rest is simple geometry. This matching problem is more known as the correspondence problem, [4]. The problem is solved using a best-fit solution and this is very hard task as there are a huge number of possible solutions. A certainty map is therefore often generated, to see how reliable each pixel in the disparity map is, see Figure 3.2. The problem in matching gets more complicated for example if specular reflection appears or if corresponding pixels in the two images have different intensity values.
Three dimensional vision

Figure 3.2 Corresponding pixels in left and right images are found and disparity and certainty maps are generated. Bright pixels in the disparity map = close pixels. Bright pixels in the certainty map = certain pixels.

Several different stereo vision methods have been developed during the years to produce a satisfactory distance map. Two well-known approaches are feature- and area-based methods. Another frequently used method is the phase-based method, described in [5], which uses the phase difference between two local filter responses to calculate the disparity. One thing the methods got in common is that the images must have a lot of texture to perform successful matching. Consequently, the presence of lines and edges are vital to achieve a good result.
3.1.3 Feature-based stereo vision
In feature-based algorithms, the intensity data is first converted to a set of features assumed to be more stable image property than raw intensities. The matching stage operates only on these extracted features. Most areas of the images will end up in the “no feature present”-class and this will lead to a great data reduction and consequently speed up the calculations. A drawback is that a dense disparity map will not be generated, because areas without features will not be considered in the matching process. To get a dense disparity map, interpolation is required and this may very well be done using normalised averaging, see section 3.2.1.

3.1.4 Area-based stereo vision
Area-based, or correlation-based [4], methods compare directly intensity values within small regions of the left and right image. In these methods one image of the image pair is selected as a reference image. For each pixel in the reference image a patch, including surrounding pixels, is chosen in order to find the same patch in the other image and hereby get the position of the corresponding pixel. The pixel with greatest similarity, in the least-square mean, is considered as the corresponding pixel. To produce certain values the matching regions should be large, but the calculation time will increase as the region enlarges. Different spatial scales, i.e. different resolutions of right and left picture, can be used to speed up the process (by starting with the lowest resolution). This is called hierarchical matching [5].
3.2 Using stereo vision in this application

Two digital cameras and the special camera stand, described in section 1.5.5, were used to produce the image pairs. The distance between the cameras was approximately 10 cm, because this gave the best result. Disparity images were then generated by an online-calculation algorithm [6] developed by Henkel. This stereo method is based on the coherence-detection between simple disparity units. An image, with the certainty for each pixel in the disparity image was also produced. Normalised averaging was used to produce a more smooth the disparity image, see Figure 3.4.

3.2.1 Normalised averaging

Normalised averaging is a filtering technique [5], which requires the certainty of all the samples. It is a special case of normalised convolution [7] and [8]. The normalised averaging formula for scalar data is:

\[ U_N = \frac{a \times cg}{a \times c} \]

\[ \text{Equation 3-7} \]

where:

- \( a \) = applicability function
- \( c \) = certainty
- \( g \) = scalar data (samples)

Improved data is produced by interpolation in small regions. More certain data gets greater weight. The normalised convolution is here used as an image enhancement method, where the samples are greyscale pixel values. Here, \( cg \) is the certainty value times they greyscale value for every pixel.

3.2.2 Choice of applicability function

The 2D-applicability function, \( a \), decides the size of the region and how much every pixel in this region should be weighted. In this test a normalised “gausslike” function, displayed in Figure 3.3, was used. The formula for the original gaussian function is:

\[ \text{gauss} = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{x^2 + y^2}{2\sigma^2}}, |x^2 + y^2| < R_{\text{max}} \]

\[ \text{Equation 3-8} \]
The 2D gauss filter can be separated into two 1D-filters, which makes the filtering more efficient. Because of the smooth shape of the filter function, see Figure 3.4, the disparity map will be low-pass filtered.

![Figure 3.3 A "Gausslike" applicability function of size 49 x 49](image)

By using the certainty values and a filter function like this, the closest neighbours, to the pixel of current interest, with great certainty will affect the most. A big filter function is used here, and hereby many surrounding pixels will affect the result.

![Figure 3.4 A smooth disparity map is made by using normalised average.](image)
Three dimensional vision

3.2.3 Results
So, the disparity data contains shape information in three dimensions and here are some results from the three main classes during ideal conditions, see Figure 3.5, Figure 3.6 and Figure 3.7. Obviously, these three cases produces very dissimilar disparity maps, and are therefore easy to separate from each other.

Figure 3.5 The left image is one image of the pair of a RFBS that generated the right disparity map.

Figure 3.6 The left image is one image of the pair of a person that generated the right disparity map.
Three dimensional vision

3.2.4 Post-processing of the disparity map

Simple and general image processing can be made on these disparity maps to distinguish them. Depth segmentation of the 3D shapes is an easy way to do this. In Figure 3.8 results from two different disparity thresholds, 150 respectively 185, are viewed.

Figure 3.8 The first column shows slices in the RFBS case, the second shows empty case and the third person case. The first row shows slices only with the closest pixels. The second row shows slices further back.
3.2.5 Conclusions stereo vision

The situations in this test were very ideal and, of course, more tests during different conditions have to be done. For example, during different lightning conditions when persons of different sizes have different positions and poses in the seat and also using different child seats.

The greatest advantages of this method are:

- It might be used as a stand-alone method, since it can distinguish between all the three occupant classes.
- It is robust to changes in lightning, as the image pair is taken simultaneously.

Some drawbacks are:

- It has problems with non-textured regions. Texture information is very dependent to what dynamic range the cameras have and the illumination of the scene.
- A satisfactory depth map can not be produced if there are a lack of structures in the scene.
- Two cameras and accurate calibration are needed. There are self-calibration techniques existing for recovering the geometry between the camera. This technique is not suitable in this case since the camera must be able to move and this equipment would add costs to the system. A static object in the scene is also required. This technique is most suitable if a camera is mounted on a mobile robot.
- The method is complex and slow.
3.3 Structured Light

While previously discussed stereo also is known as passive stereo vision, structured light is a so called active stereo vision method, which is performed using active illumination. Like stereo methods the result of this method is a depth map. This method only requires one camera but needs special lightning equipment. The scene is illuminated by light patterns, for example grids or stripes, and an image is taken. The light can for example be generated by a laser-scanner. Using the deformation caused by the object, the shape of the scene can easily be extracted from that image. By knowing camera and projector geometry the depth can be calculated by triangulation.

![Diagram of structured light](image)

*Figure 3.9 A scene is illuminated with a grid pattern and an image is taken.*

The calculations in this method are very simple and therefore the method is fast. But the accuracy and resolution in depth is low and only a coarse depth is generated.

3.3.1 Conclusions structured light

In this method, quite well controlled lighting is required and the main target is to extract the pattern that is projected on the scene. As long as this can be done it will be possible to calculate the distance map. The correspondence problem is totally absent and it is also faster than stereo, but cannot give the same resolution in the distance map.
Three dimensional vision

Some other advantages of this method are:
• It is robust to rapid light changes, sweeping shadows etc. as long as they are controlled.
• It deals with scenes that do not contain sufficient features, such as edges or corners, which are associated with intensity discontinuities, for the stereo matching process.

Its shortcomings are:
• The system must be pre-calibrated.
• A special light pattern is needed, besides a camera.
• It might be dangerous to expose people to laser or other light sources.
• If the light is absorbed or can not be extracted for other reasons, this method will fail.
Three dimensional vision

3.4 Conclusions using 3D features
By using three dimensions great robustness will be obtained. It will not be easy to fool, because of the introduction of one more dimension and robustness to light changes is also achieved by using stereo vision or structured light. The main drawback is that extra equipment is needed.

3.4.1 Classes that possibly can be detected
- Person
- Empty seat
- Child seat
4 Motion

“A lot of information can be extracted from time varying sequences of images, often more easily than from static images. For example, camouflaged objects are only easily seen when they move. Moreover, the relative sizes and position of objects are more easily determined when the objects move. Even simple image differencing provides an edge detector for the silhouettes of texture-free objects moving over any static background.”

Robyn Owens, Department of Computer Science, The University of Western Australia [2]

There is one source of visual information that probably is used by all biological vision systems, and that is motion information. It seems like humans are skilful at identifying things in this way. Consequently, motion is vital for us to interact with our surroundings. For example, driving a car in traffic would practically be impossible if we did not have the talent to detect moving objects like cars and other vehicles. To distinguishing between shapes in still images can be very difficult, but if motion is used as an additional cue then more information is available.

Let us use the same example as in the 3D case. Assume a method has the purpose to detect faces of humans in a 3D scene. Like in 3D vision an extra dimension, motion, is introduced. By using this extra information and the fact that a person nearly always moves slightly, the ability to distinguish between a real human face and an image of a face is obtained. Segmentation by motion is a widely used approach in computer vision and is an effective way to extract objects in a scene.
4.1 Optical flow

The optical flow measures differences between images in an image sequence. In other words, motion is detected. A velocity vector, \((u,v)\), is calculated for each pixel. Suppose that we have a volume \(f(x,y,t)\) of an object. This object, which is a part of the image, is moving with a translation in the xy-plane. A small distance and a small time later:

\[
f(x, y, t) = f(x + \Delta x, y + \Delta y, t + \Delta t) \quad \text{Equation 4-1}
\]

![Figure 4.1 An object which moves with a translation in the xy-plane during the time \(\Delta t\).](image)

Taylors formula gives:

\[
f(x + \Delta x, y + \Delta y, t + \Delta t) = f(x, y, t) + \Delta x \cdot \frac{\partial f}{\partial x} + \Delta y \cdot \frac{\partial f}{\partial y} + \Delta t \frac{\partial f}{\partial t} + ... \quad \text{Equation 4-2}
\]

If higher order terms in Equation 4-2 are neglected, Equation 4-1 gives

\[
\Delta x \cdot \frac{\partial f}{\partial x} + \Delta y \cdot \frac{\partial f}{\partial y} + \Delta t \frac{\partial f}{\partial t} = 0 \quad \text{Equation 4-3}
\]
Let $u = \Delta x/\Delta t$, $v = \Delta y/\Delta t$ then:

$$u \frac{\partial f}{\partial x} + v \frac{\partial f}{\partial y} = \frac{\partial f}{\partial t}$$

Equation 4-4

This equation is more known as the optical flow constraint equation, where $(u,v)$ is the optical flow.

### 4.1.1 The aperture problem

If the local image curvature is one-dimensional, can only one component of the $(u,v)$-vector be calculated. This problem is often called the aperture problem, which is illustrated in Figure 4.2.

---

**Figure 4.2** Illustration of the aperture problem. The right box only moves in the v-direction and the left box is moving in both directions. The true flow can only be found in the neighbourhoods of the corners and near the top and bottom of the right box.
4.1.2 Optical flow using second derivatives

The optical flow constraint equation itself is not sufficient to calculate the optical flow. There are two unknowns, \( u \) and \( v \), and to solve the constraint equation additional constraints must be added. A lot of methods with different constraints have been proposed to calculate the optical flow. Some of them are discussed in [9]. The simplest way to derive the optical flow is done using difference-based techniques. We now look into such a method based on second derivatives. This method uses the assumption that \( u(x,y) \) and \( v(x,y) \) varies slowly in the image. The derivatives of \( u \) and \( v \) are therefore neglected, i.e. the acceleration for each pixel is set to zero:

\[
\begin{align*}
    u_x = u_y = v_x = v_y = 0
\end{align*}
\]

Equation 4-5

With this assumption the \( x \)- and \( y \)-derivatives of Equation 4-4 are:

\[
\begin{align*}
    u \cdot \frac{\partial^2 f}{\partial x^2} + v \cdot \frac{\partial^2 f}{\partial x \partial y} & = -\frac{\partial^2 f}{\partial x \partial t} \quad \text{Equation 4-6} \\
    u \cdot \frac{\partial^2 f}{\partial x \partial y} + v \cdot \frac{\partial^2 f}{\partial y^2} & = -\frac{\partial^2 f}{\partial y \partial t} \quad \text{Equation 4-7}
\end{align*}
\]

Now \( u \) and \( v \) can be calculated:

\[
\begin{align*}
    u & = \left( \frac{\partial^2 f}{\partial x \partial t} \cdot \frac{\partial^2 f}{\partial y^2} \right) - \left( \frac{\partial^2 f}{\partial x \partial y} \cdot \frac{\partial^2 f}{\partial x^2} \right) \\
    v & = \left( \frac{\partial^2 f}{\partial y \partial t} \cdot \frac{\partial^2 f}{\partial x^2} \right) - \left( \frac{\partial^2 f}{\partial x \partial y} \cdot \frac{\partial^2 f}{\partial y^2} \right) \quad \text{Equation 4-8, 4-9}
\end{align*}
\]

One way of implementing this method using separated convolution kernels is described in [10].
4.2 Using optical flow in this application

A mitsubishi artificial retina, see section 1.5.5, generated the test sequences, and the optical flow was calculated using an implementation of the second derivative method done by Karl Munsin. This is fast and rather good compared to other implementations, see [11] for evaluation and detailed description of different techniques. In Figure 4.3 some images frames from test sequences are shown. Movements, larger than one pixel, between five consecutive images were detected. These areas are coloured in the image.

![Figure 4.3](image)

*Figure 4.3 Top left: Empty, no motion. Top right: RFBF, movement of the left arm (Simulation with puppet). Bottom left: Person, adult that moves the head slightly. Bottom right: Person, child that moves the head and the left shoulder (Simulation with puppet).*
The colours correspond to different movement directions, which are defined in Figure 4.4. Motion segmentation can be done more easily if the directions of the movements are used.

![Figure 4.4 Colour coded movement directions.](image)

By using constraints related to shape, size, position etc. classifications of different objects may be performed. One problem with this method is if objects sit still for a while. By using older estimates the certainty of new estimates can be increased and by accumulate motion data a more dense motion field can be produced. A procedure which attempts to integrate and update optical flow estimates from multiple frames, called accumulation of evidence, are discussed in [9].

4.2.1 Conclusions optical flow

The test shows the potential of using optical flow to segment moving objects like persons. The test shows that lines and edges are vital to be able to estimate motion. Therefore, good images are needed to be able to detect motion, and consequently the dynamic range of the camera has to be high. The method is less sensitive to light changes that direct differing techniques, but still not quite. Sweeping shadows will for example be detected as a moving object. The method is also very slow.
4.3 Conclusions using motion

Methods depending on motion are not likely to be used as a stand-alone method, because only things that are moving can be detected. People that are paralysed or sleeping do not move much and infants in a rear RFBS are often covered with blankets and almost no motion will be able to be detected. Therefore, this method cannot solve the problem alone. However, motion will probably be a good complement to other methods to detect persons and make it more probable that the seat is empty. How small movement that can be detected is depending on the resolution of the camera. Most motion methods provides poor robustness to rapid light variations in daylight, sweeping shadows etc because this will be detected as motion.

Some good things with motion are:
- Only one camera is needed.
- The fact that a person always moves can be used.

4.3.1 Classes that possibly can be detected
- Person
"Color information is useful for recognition, but the measured image color of surfaces depends on the scene illumination. Human vision exhibits color constancy as the ability to perceive stable surface colors for a fixed object under a wide range of illumination conditions and scene configurations. A similar ability is required if computer vision systems are to recognize objects in uncontrolled environments."

G. Healey.

Humans can perceive colours in the light spectrum between the wavelengths 400 and 700 nm. The retina has two types of sensors, rods and cones. The rods, which sense the intensity of light, are used for night vision and the cones are used for daylight vision. There are three types of cones, each sensitive to different wavelengths, namely red, green and blue. There are many colour models available, which describes and measures the colour, but no finite set of colours can be combined to display all possible colours.

5.1 Colour models

Like the human system most colour models consists of three primary colours. In the 1931 the XYZ model was defined as a standard by CIE (Commission Internationale de l’Eclairage). This model has the three non-physical primaries X, Y and Z, which only exists in a mathematical sense.
The model is additive and by combining different amounts of these colours a certain colour \( C \) can be made:

\[
C = XX + YY + ZZ \quad \text{Equation 5-1}
\]

This model is further described in [5]. Other more intuitive colour models have also been defined to fill certain purposes.

5.1.1 The RGB colour model

RGB is the most widely used model in computer hardware and cameras and, similar to our eyes, this model represents a colour as the three independent components red, green and blue. Like the XYZ system, RGB is an additive model and combination of \( R \), \( G \) and \( B \) values generates a specific colour, \( C \):

\[
C = RR + GG + BB \quad \text{Equation 5-2}
\]

This model is often represented by a 3D-box, with \( R \), \( G \) and \( B \) axes, see Figure 5.1.

![Figure 5.1 The RGB colour model unit box.](image-url)
The corners of the box, on the axes, correspond to the primary colours. Black is positioned in the origin, (0, 0, 0) and white in the opposite corner of the box, at (1, 1, 1) and is the sum of the primary colours. The other corners represents combinations of two primary colours, for example adding red and blue gives magenta, (1, 0, 1). Shades of grey are positioned long the diagonal from black to white. Still, this model is hard to comprehend for a human observer, because the human way of understanding and describing colour is not based on combinations of red, green and blue.

5.1.2 The HSV colour model

HSV is a colour model that is more intuitive to humans. To specify a colour, one colour is chosen and amounts of black and white is added, which gives different shades, tints and tones. The colour parameters here are called hue, saturation and value. In a three-dimensional representation, see Figure 5.2, hue is the colour and is represented as an angle between 0° and 360°. The saturation varies from 0 to 1 and is the purity of the colour, for example is a pale colour like pink less pure than red. Value varies from 0 at the apex of the cone, which stands for black, to 1 at the top, where the colours have their maximum intensity.

![Figure 5.2 The HSV colour model cone.](image)
5.2 Segmentation by skin colour

Colour is a useful attribute to human vision, especially in the ability to recognise and sort out objects in the surrounding world. Likewise, colour has also been shown to be a useful feature in computer vision. Colour can be used in order to segment, recognise and classify objects and areas in images. As for other computer vision methods that are to be used in more or less unconstrained scenes, invariance to changing lightning conditions is an important issue. Great efforts have already been done, especially to find human skin colour, and new proposals are constantly presented in this area.

5.2.1 Skin colour segmentation using HSV space

Earlier studies have shown that all kinds of human skin, no matter the race, are gathered in a relatively small cluster in a suitable colour space. In [12], Garcia and Tziritas states that direct access to Hue mainly encode skin colours. Tsekeridou and Pitas showed, in [13], that human skin-colours are positioned in a small cluster of the HSV space. Their suggestions to appropriate thresholds for hue saturation and value values were defined like this:

\[
\begin{align*}
0^\circ \leq H &\leq 25^\circ, \quad 335^\circ \leq H \leq 360^\circ \\
0.2 \leq S &\leq 0.6 \\
V &\geq 0.4
\end{align*}
\]

By using hue and saturation great robustness to lightning changes is obtained, because different lightning do not affect these parameters much. Since most cameras produces RGB pixels, a conversion to HSV has to be done first, see [14] for details. The RGB values are camera dependent because different cameras produce different values. Therefore, the thresholds must be adjusted to fit the specific camera that is used.
5.2.2 Skin colour segmentation using other spaces

Many other colour spaces have been proposed during the latest years in order to segment skin colour areas. Here are some examples:

- In [15], Wu, Chen and Yachida detects faces using both skin and hair colour. First they convert RGB colour information to CIE’s XYZ colour system and then converting this information to the UCS colour system. UCS colour representation is similar to the sensitivity of human eyes.

- In [12] Garcia and Tziritas uses both approximations of the YCbCr and HSV skin colour subspaces to provide quantised skin colour regions.

- Schumeyer and Barner suggests in [16] that L*a*b* is a suitable colour space in this application.

- In [17] skin colour is modelled based on a particular reflectance model of the skin, pre-knowledge about camera parameters and the spectrum of the light sources that are used. Fluorescent lamps with different correlated colour temperatures (CCT’s) were used and the skin colour area in the chromaticity plane for the different CCT’s were calculated. They showed that the localisation of the skin colour area changes drastically in the chromaticity plane, as the colour temperature of the lightning changes.

- In [18] skin detection is performed using a skin filter, which relies on Hue and Saturation, but also texture information. The face detection is performed on a greyscale image containing only the detected skin areas.
5.3 Using HS segmentation in this application

A small set of images with one empty seat and some faces of different persons (all Caucasians except one Arabian) were generated in a car and are viewed in Figure 5.3.

![Image of faces](image1.jpg)

*Figure 5.3 The image face set.*

The digital camera described in section 1.5.5 produced these images. Different operations were performed in Matlab to generate the images in this chapter.

5.3.1 Hue segmentation

In this test, a small part of the hue scale were used and the threshold were:

\[ 0^\circ \leq H \leq 25^\circ \]
This hue cluster is shown below in Figure 5.4.

Figure 5.4 This hue cluster was used for segmentation.

By using this cluster the corresponding images shown in Figure 5.5 were produced. In this binary image the white areas are considered as possible skin regions. This example shows that hue mainly encodes skin colour.

Figure 5.5 The skin colour regions using only hue.
5.3.2 Saturation segmentation
The following saturation threshold was used to perform the result shown in Figure 5.6:

\[ 0.17 \leq S \leq 0.8 \]

Figure 5.6 The skin colour regions using only saturation.

5.3.3 Hue and saturation segmentation
By using thresholds for both hue and saturation some further false areas can be discarded, see Figure 5.7.
5.3.4 Post-processing of skin colour areas

Different filtering techniques can be applied on the binary images in Figure 5.7 to get more homogeneous skin colour areas. For example, a median filter technique, described in [10], can be used to get rid of single noise pixels. In this technique, the output is set to the median (not the average) pixel value in the neighbourhood. If a median filter of size 15*15 is applied on the skin colour image in Figure 5.7, the result, viewed in Figure 5.8, is received. According to P-E Danielsson in [10], an amplification of the noise suppression and edge sharpening might be obtained by iterative median filtering. So, this might be explored further.

Another technique, which also reduces noise pixels, can be done using binary erosion by some pixels followed by dilation by the same amount of pixels. Assume that the objects are white and have pixel value 1, and the background is black and has pixel value 0.
Colour vision

Erosion is performed by, within a certain distance \( r \), setting all the pixels on the object border to zero (black). Dilation is performed by, inside a certain distance \( r \) from the object, setting all the background pixels to 1 (white).

![Median filtered images. Filter size = 15*15.](image)

By introducing further constraints greater robustness will be obtained. Such constraints could for example be related to shape and size of the face, for example:

- The face has an elliptical shape.
- In this test, the faces in the images are larger than 40*40 pixels.

So, by using pre-knowledge it will be easier to localise the face and discard false areas.
5.3.5 Using HS segmentation as a data reduction

A procedure presented in [18] can be used to perform more reliable face detection. Here, the binary skin map is multiplied by the original image and further face detecting methods, see chapter 6, can be applied only on the skin colour regions. Hereby, the search area is drastically reduced. Before the multiplication, the face areas might be expanded to ensure that most part of the face would be visible. This is performed using dilation (8-connective), which briefly are described in section 5.3.4. In this example the distance \( r = 5 \) was used and the result of this operation is shown in Figure 5.9.

![Figure 5.9 Expanded face regions.](image)
After this operation, the expanded image is multiplied by the original image, see Figure 5.10, and further methods can be applied after this data reduction.

![Figure 5.10 Image ready for further processing.]

5.3.6 Conclusions using HS segmentation

So, simple skin colour segmentation may easily be done on colour images by using a small cluster of the hue and saturation space and by using simple binary operations face detection may be performed. By using the approach suggested in this chapter these favourable characteristics will be obtained:

- Great independence to changing lightning conditions.
- Robustness to orientation. The rotation of the face will not matter much.
- The method is very fast.
- Information like height of the person can easily be determined.

Problems will be introduced if objects in the background end up in the skin colour space and, as can be seen in the last row of Figure 5.5, two sweatshirts ends up in the skin colour cluster. So, further processing has to deal with such problems. Additionally, the method does of course not
Colour vision

manage to find the skin regions if they are hidden behind something like disguises, newspapers etc or if the head is rotated too much, for example if a person is facing the back seat.

To get a robust segmentation as possible more colour spaces can be explored. A much larger set of face images, containing all sort of skin colours during different lightning situations, must also be used to find appropriate thresholds and hereby define an accurate skin colour cluster.
5.4 Conclusions using colour vision

These characteristics are general for colour vision methods:

- Great robustness to rapid light variations in daylight, sweeping shadows can be achieved by using suitable colour spaces.
- More information is obtainable.
- Of course, a colour camera is required. The camera may be slower than a greyscale camera, because the data will increase three times. A colour camera may also be more expensive.
- The major problem to deal with is that colour information will disappear if it is too dark. Therefore, special lightning or backup methods in the dark cases are required.

It should be noted again that different cameras give different colours values under the same conditions and therefore the specific camera that is to be used has to be evaluated.

5.4.1 Classes that possibly can be detected

- Person
6 Face detection

“Face recognition is not so much about face recognition at all - it is much more about face detection! It is my (and many others’) strong believe, that the prior step to face recognition, the accurate detection of human faces in arbitrary scenes, is the most important process involved. When faces could be detected exactly in any scene, the recognition step afterwards would not be so complicated.”

Robert Frischholz

We will in this chapter take a brief look into some common face detection methods, which in other applications mostly are used as a first approximate localisation of faces in images. In those cases, this is done in order to reduce the search area for further and more accurate (and often computationally expensive) facial feature detection methods.

In recent years, great efforts have been made with the aim to develop methods that detects faces in images. The fields of application are huge, for example person identification in security systems, human computer interaction and video telephone. Face detection is the first, and the most difficult, step in a fully automatic human face recognition system. In fact, this is one of the harder problems in pattern recognition.
Face detection

There are several problems to deal with in face detection, due to face variations like:

- scale
- pose, orientation
- facial expression
- brightness, caused by varying illumination
- disguise, for example glasses and beard

Here, some different ways to find faces in images will shortly be reviewed. The choice of approach must be based on what camera resolution that is available. High resolution is needed to find features such as eyes, mouth and eyebrows, an example can be found in [19]. The next sections will treat methods, which do not require very high resolution images.

6.1 Finding a face by motion

Motion-based face detection methods uses the fact that a person moves almost all the time. This method is very useful if the only moving object in the scene is the person. So, if the background is constant, it is easy to find the head once the motion of the person is detected. Motion is probably a very useful attribute and will increase robustness to variations in the face like beard, glasses, race, masquerade attributes, camouflage etc. One method to detect motion is optical flow, see section 4.1. One work using motion in face detection is presented in [20].

6.2 Finding a face using a set of examples

Example-based, or template-based face detection uses a set of examples of faces and the task is to find objects similar to these, which hopefully also are faces. So, this is merely a matching problem where the aim is to compute distance measures between the example set and every possible region in the image. Hereby, this method is very computationally expensive. The choice of examples requires great efforts, because the result of matching is highly dependent on how large the example set is and how good the set represent all possible faces. Of course the computation time will increase as the example set get larger. There are many different ways to perform pattern matching and some approaches are described in [10]. In [21] Sun and Wei presents an eigenface method for face recognition. In this approach some face images are decomposed into a small set of characteristic feature images, so called eigenfaces. The comparison is performed by projecting a new face into the low-dimensional linear face space, defined by the eigenfaces.
Face detection

6.3 Finding a face using neural network

Instead of writing own computer programmes and criteria’s a neural network can be trained to detect faces (See chapter 7 for more details on neural networks). Such a system is presented in [22], and works as following:

“A retinally connected neural network examines small window of an image, and decides whether each window contains a face. The system arbitrates between multiple networks to improve performance over a single network” [22]

An interactive web demonstration can be found on this web page http://www.ius.cs.cmu.edu/demos/facedemo.html. This system takes an arbitrary greyscale image and tries to find faces in it. The performance of this system is quite well, but non-frontal faces are hardly detected and the approach is extremely computationally expensive. Additionally, very time consuming training was needed to achieve the result. They also found it hard to characterise non-face representative images, but in the car application this may be a bit easier. See section 7.4 for further conclusions.

6.4 Finding a face by colour

This area is earlier discussed in chapter 5.
6.5 Using face detection in this application

See section 5.3 for examples and results from a colour segmentation method.

6.6 Conclusions using face detection

A face detection method cannot be used as a “stand alone” method, but if a face is found and located in the images a lot of conclusions can be made. It will also be more probable that the seat is empty if no face is found. As mentioned there are many different ways to detect a face and the computation speed depends of the choice of algorithm. Some general conclusions are that:

- A lot of research has been made in this area recently and can be studied to get new ideas that may be suitable in car application.
- The methods are not robust to disguises and obstruction.

Some problems in face detection can be limited, because of the quite controlled environment in the car. Here are some restrictions that can be introduced:

- The size and resolution of the images are always the same.
- The size of a face does not vary as much as in the unconstrained case.
- Persons are normally looking straight ahead, and a camera can hereby capture the face most of the time.
- The positions of the head do not vary so much in time and the face is most of the time inside a certain area.

6.6.1 Classes that possibly can be detected

- Person
7 Neural networks

“A whole new science was born with the aim of producing such intelligent machines - the subject of artificial intelligence or AI. In fact that has not happened - initial efforts to create computers with mind-like reasoning have failed miserably. Many researchers now believe that part of the reason for this failure was that traditional computers function in a way very different from the brain and that the key to true intelligent machines lies in understanding in detail the functioning of the brain and emulating this with artificial neural networks.”

[23]

The earlier described methods in this thesis rely on the programmer’s ability to produce reliable algorithms. There are other ways to perform pattern recognition and to make classification. This chapter contains a short introduction to neural networks.

Like the human brain a neural network, or more properly artificial neural network, is learned to make the right decision. The human brain contains many billions nerve cells, called neurons. Each neuron is physically connected to many other neurons and together they form a complicated intercommunicating network. So far, human and other biological neural networks are much more complex than artificial neural networks.
7.1 Characteristics

From now on we refer neural networks to artificial neural networks. A neural network is a network of nodes and each node is a simplified model of a real neuron. These nodes are connected to communication channels, so called connections. To learn in a proper way, the network needs a lot of suitable training data. Together with the input the corresponding output decision has to be brought.

An ordinary computer works sequentially and is a machine, not a mind. Neural networks works without conventional programming. The network is trained, not programmed. They are able to solve problems that are hard to or cannot be programmed. Training a network may in some cases be faster than writing traditional software. Some other favourable characteristics for the neural networks are that they can find relationships in data that are unknown for humans, and that they have great potential for parallelism since the computations of the components are largely independent of each other.

![Figure 7.1 A three-layered neural network.](image-url)
Neural networks

7.2 Back propagation networks
A common and often useful network is the type that learns by back propagation. Typically the nodes are arranged in layers and pass data through the connections. Each connection has an associated adjustable value, which are called weight. With suitable weights, such a network can model any computable function. In the training each weight is adjusted to reduce errors. Figure 7.1 shows a three-layered network, further described in [24], that learns by back propagation.

7.3 Using neural networks in this application
An input to a neural network can be any data, for example image pixel data. Therefore, a neural network can be trained to recognise different patterns, like faces or other features. One method that detects faces is presented in [22] and is also shortly reviewed in section 6.3. More useful information about neural networks and how to train them is presented in [24].

7.4 Conclusions using neural networks
There are software programmes on the market that generates neural networks so the main task will be to generate sufficiently good training data and choose appropriate features as input. Neural networks are suitable when:

- An algorithmic solution can not be formulated.
- A lot of training data is available or can be produced, in order to make it possible to obtain a robust network.

Some drawbacks are following:

- A neural network can be considered as a “Black box”. This is a drawback if you want to know on what basis a certain decision is made. So, the decision is taken entirely by the machine.
- Can be trained wrong, and this will maybe not be detected immediately.
- Neural networks are often computationally expensive especially if there are a lot of data.
- It takes a lot of time to train.

7.4.1 Classes that possibly can be detected
- Person
- Empty seat
- Child seat
8 Evaluation

In this chapter the earlier discussed methods in this thesis will be evaluated, see Table 8.1. A comparison to other passenger detection systems will also be presented.

8.1 Evaluation of the computer vision methods

The criteria, which are used in this evaluation, are following:

- Robustness to fast and slow changes in lightning.
- What result will be performed and how reliable is the information and how easy is it to use and interpret.
- Is the method a stand-alone method, i.e. can all occupant classes be detected using only this method.
- How complicated it is to implement the algorithm.
- Whether calibration, extra equipment or training is needed.
- As stated in the introduction section 1.5.4, the speed is not very important in the classification. Although, we will look at this because it is preferable to produce fast results. An optical system may also be used for real-time tracking during a crash in order to measure distance between occupant and the airbag, which requires fast methods.
- If the information always can be obtained, or under which circumstances.

The conclusions from the methods in this thesis and also from the earlier work by Autoliv SAGEM in France [26] are presented in Table 8.1.
### Evaluation

**Table 8.1 Evaluation of the computer vision methods**

<table>
<thead>
<tr>
<th>METHOD</th>
<th>Advantages</th>
<th>Drawbacks</th>
<th>Classes</th>
<th>Detects</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Structured light</td>
<td>Gives 3D-info. Robust to light changes. Fast and simple.</td>
<td>Cannot produce a dense disparity map. A special light pattern is needed.</td>
<td>Person Empty RFBS</td>
<td>3D shape Depth Distance</td>
<td>Good! Stand-alone!</td>
</tr>
<tr>
<td>Optical flow</td>
<td>Easy object segmentation. No special gear is required.</td>
<td>Slow and not very robust to light changes. Cannot detect non-moving objects.</td>
<td>Person Movement 2D shape</td>
<td>Movement 2D shape</td>
<td>Slow, but good to use motion as a complement.</td>
</tr>
<tr>
<td>Lines (Autoliv SAGEM, France)</td>
<td>Fast and simple. No special gear is required.</td>
<td>Poor robustness to changes in lightning. Requires very good lightning. Hard to interpret line images.</td>
<td>Empty RFBS Lines</td>
<td>Empty RFBS Lines</td>
<td>Poor.</td>
</tr>
<tr>
<td>Face NN</td>
<td>Requires no programming.</td>
<td>Training takes quite a while. &quot;Black box&quot; and slow. Requires a lot of training data.</td>
<td>Person Faces</td>
<td>Faces</td>
<td>Slow and needs training.</td>
</tr>
<tr>
<td>Face Motion</td>
<td>Easy object segmentation. No special gear is required.</td>
<td>Not robust to light changes. Cannot detect non-moving objects. Speed depends on the motion method.</td>
<td>Person Faces</td>
<td>Faces</td>
<td>Good complement method.</td>
</tr>
<tr>
<td>Face Example faces</td>
<td>Speed depends on number of example faces. Needs a lot of programming. Needs a lot of example faces.</td>
<td></td>
<td>Person Faces</td>
<td>Faces</td>
<td>Not so good.</td>
</tr>
</tbody>
</table>

58
8.2 Optical system vs. ultra sonic and weight

So, what is better with an optical system preferred to the other present occupant sensing system and what drawbacks are there?

8.2.1 Weight system

Autoliv’s current weight system, called OWS (Occupant Weight System), is viewed in Figure 8.1 performs classification based on:

- the weight of object and
- the centre of gravity (COG) for the object (the COG gives information about how the object is positioned in the seat).

![Figure 8.1 Ows with four load cells, processor and a buckle switch.](image)

8.2.2 Ultra sonic system

The present ultra sonic system, called OSS (Occupant Spatial Sensing) has the following classes:

- out-of-position object
- in position object

The decision-maker is a neural network, which has been trained to recognise occupants due to these classes.
8.2.3 Optical system

Some advantages of an optical system (with cameras) are following:

- More precise classification is possible to be made (not only classification in the three main classes: person, empty seat and RFBS). For example can the height of person be determined.
- No training is needed and the decisions are based on algorithms and not made by a neural network, as in the present OSS system.
- A camera gives much more information than the other systems and can distinguish between living and dead objects.
- The cameras should be positioned in the roof, where it not will be damaged or obstructed so easy. The OWS load cells are positioned under the seat and may more often be exposed to moisture, hits and changing temperatures.
- Further features can be used. For example can a real-time high-resolution distance measure, between the occupant and the airbag, be made. The camera can record post-crash information and be sent to report injuries. Other fancy features as information to an automatic neck support adjuster can be produced.

Some drawbacks with a camera system:

- More data is available, and needs more processor power.
- It has to cope with both fast and slow varying illumination.
- The occupant can be obstructed and moisture can appear on the lens.
This work shows that there is a great potential for an optical occupant sensing system in order to detect and classify occupants, but certainly much work has to be done before such a system are able to go into mass production. The choice of methods to use will not only be based on how the method perform, for instance it will also depend on:

- Cameras, which are to be used
- Available processor power
- Law requirements
- Autoliv customer and manufacturer requirements
- What other features the optical system are supposed to have
- Costs
- The position of the cameras

A 100% functional system is probably impossible to achieve, but should of course be the aim. If more than one camera is used greater robustness may be achieved. A suggestion of how three cameras for each seat can be placed is shown in Figure 9.1 (all the cameras are placed in the roof of the car).

Figure 9.1 Overview of the suggested camera placement.
Discussion

Using more than one camera makes it possible to perform classification even if one camera is obstructed. If each camera uses independent methods for classification and gives the same result a higher robustness will be obtained. This set-up is suitable for an automatic occupant detecting and tracking system, where the distance between the persons face and the airbag unit is calculated by the side camera and the front cameras are able to perform 3D classification using stereo.

9.1 Further improvements

A patent pending technique developed by M. Klomark, A. Olsson and K. Munsin called ABE (Active background elimination) [25], which perform object extraction and data reduction will probably make it possible to perform rough stereo disparity data at sufficiently high speed maps and real-time distance calculations (head to airbag).

Some improvements may also be made if pre-knowledge is introduced. Knowledge about the geometry can be used. The distance to the object will, for example, always be less than one meter from the camera. Reference images may also be used to compare with and thereby be able to extract objects. However, the choice of reference image is non-trivial.

Another possible approach is to enable users to register personal data as weight, length, age etc. Hereby, will this information be accessible when the user “logs on” to the system before the trip starts.

Additional information will probably be available to the main airbag controller such as:
- Buckle switch signal.
- Ignition and door switch signals, which may be used as start-up signals.
- Weight information from a weight system, which may be used to make more robust classification.
10 Summary

The purpose of this thesis was to study the possibility to use computer vision methods to detect and classify objects in the front passenger seat. This work presents different approaches to solve this problem and evaluates the usefulness of each technique.

The work shows that computer vision has a great potential in order to provide data, which may be used to perform reliable occupant classification. The main conclusion is that methods, which provide robustness to fast and slow variations in lightning, are most suitable in this application. Future choice of method to use depends on many factors, for example costs and requirements on the system from laws and car manufacturers. Further evaluation and tests of the methods in this thesis, other methods, the ABE approach and post-processing of the results should also be made before the final classification algorithm may be written.

A combining system is preferable to use, because every method has problems in different situations and if two or more methods are used they may complement each other. Further, a more reliable decision can be made if different methods give the same result.
Bibliography


Bibliography


http://www.cs.cmu.edu/afs/cs.cmu.edu/user/har/Web/usr2/CMU-CS-95-158R/www-techreport2.html


Appendix

Moore’s law

Initially this was an observation made in 1965 by Gordon Moore, co-founder of Intel, that the number of transistors per square inch on integrated circuits had doubled every year since the integrated circuit was invented. Moore predicted that this trend would continue for the foreseeable future. In subsequent years, the pace slowed down a bit, but data density has doubled approximately every 18 months, and this is the current definition of Moore’s Law, which Moore himself has blessed. Most experts, including Moore himself, expect Moore’s Law to hold until 2020. By then the technology will be pushing the "Point One Limit." This is when creating chips with a laser beam smaller than 0.1 micron (roughly the width of a coil of DNA) will present problems as quantum theory problems come into play.

http://www.georgehernandez.com/vocCPU/m.htm