ERASMUS Intensive Program 2001

7^{th} – 18^{th} May, Pavia, Italy

3D Vision Group

Project title

" Computing relative disparity maps

from stereo images"

Members: **Amr Ahmed, University of Surrey, UK**
Christian Cucculelli, University of Pavia, Italy
Eleni Kokkinou, University of Surrey, UK
Halima Habieb – Mammar, University INSA of Lyon, France

Supervisors : **Robert Sablatnig, Vienna University of Technology, Austria**
Martin Kampel, Vienna University of Technology, Austria

Table of contents

1. Introduction

This project is entitled "**Stereo Vision**" and it will include the work of the 3D-Vision group formed for the purposes of the **Erasmus Intensive programme in Computer Vision 2001** held in Pavia, Italy, from the 7th until the 18th of May. The members of the group represented three different universities: Institut National des Sciences Appliquées de Lyon, *France*, University of Pavia, *Italy* and University of Surrey, *UK*, under the supervision of lecturers from Vienna University of Technology, *Austria*. Former contact or collaboration between the participants was not available and, apart from the mentors of the team, the members had no previous experience in the area of 3D-Vision.

Summarising the two weeks, the team started by having introductory lectures to the field of 3D-Vision and then in the more specific topic of stereo images. Discussion followed in order to clarify further the topic of stereo images, then the aim, and the objectives of the project were defined. A literature survey was performed regarding methods of stereopsis and computing disparity. At the end of the first week, the team had become more familiar with the topic of stereo images, obtained stereo images using digital camera and presented the basic theory and the project plan to the course in a way of formal presentation. During the second week, the group performed the practical work by computing algorithms, evaluating the results and preparing a report and final presentation.

The project can be better defined by presenting its aim and objectives. Thus, the aim is to extract depth information and compute disparity maps from stereo images. The objectives of the project are: first, area defined by the window size is used to select the point positions in the left image. An area-based algorithm is utilised in order to find the corresponding point positions, which is by computing correlation between the window defined in the left image and a same size window that scanned through the right image. The points with maximum correlation value are the matching points. Finally, the depth information is extracted by evaluating disparity maps between the matching points.

The next section of the report includes basic theory information about stereo vision and a brief literature review about different algorithms. In section 3 the methodology followed is explained and the results are shown in section 4. Section 5 includes general discussion about the problems faced during the practical work, ideas for improving the used method and finally conclusions.

2. Theory

Vision –in general- is defined as the process of investigating/discovering the surrounding world from images. It describes what is present and where it is [4]. This process uses images as an input and generate 3D perception as an output which represents the surrounding world.

Computer vision is the computer implementation of the vision process that allows using computers to reconstruct 3D scenes and recover objects depths, distances,..etc. from images. Computer vision sometimes referred to as 'Machine Vision'.

2.1 Computer Vision Techniques

Most of computer vision techniques are based on 2D images. As 3D information is most commonly needed, a lot of researches have been done in how to obtain 3D information from 2D images. Generaly, computer vision techniques can be categorized to two main categories; "Active Vision", and "Passive Vision".

Active vision is referred to the techniques that use active source for energy emission such as a laser source. Light is emitted from that active source, reflected on the object's surface and detected by sensors (such as cameras). One of the common active vision techniques is the range finder where the active source (or the object) moves to scan all the object surface. It is obvious that active vision techniques have some limitations. For example, it is difficult to use it in outside measurements or wherever there are conflict between the light source and surrounding lights. Each acquisition system can be used for certain range of object's sizes. Also, the object's color may conflicts with the light source used (For example, red laser light projected on red object). Another issue is the accuracy and synchronization of movement of either the object or the active source in order to scan the whole object.

Passive vision does not use any specific active sources. Off course there should be some source of light somehow as no vision is available without light. The difference is that this light is not specific to the technique and is not directly used in the calculation algorithm (as explained below). The most common technique in this category is the "Stereo Vision". The main problem in passive vision is finding corresponding points in different images.

The basic principle used in recovering 3D information is the triangulation principle. In active vision techniques, a triangle is created between the light source, the object, and the sensor (CCD camera). In passive techniques, a triangle is created between the object and two sensors (cameras). This explain why the light source is not directly involved in calculation in passive vision.

2.2 Stereo Vision

Stereo vision means constructing 3D information from at least two slightly different 2D images. Stereo vision concept is similar to human vision system. Human vision system is based on two eyes for the aquisition of images from the world. Each eye captures the same scene producing a 2D image. However, distance between the two eyes results in a slightly different 2D images of the same object. This is a very useful phenomena that enables human to perceive 3D information of the surrounding world. Human brain is capable to use these two 2D images to generate 3D information. 3D perception is important for human being in order to have sense of distances, angles, shapes and volumes.

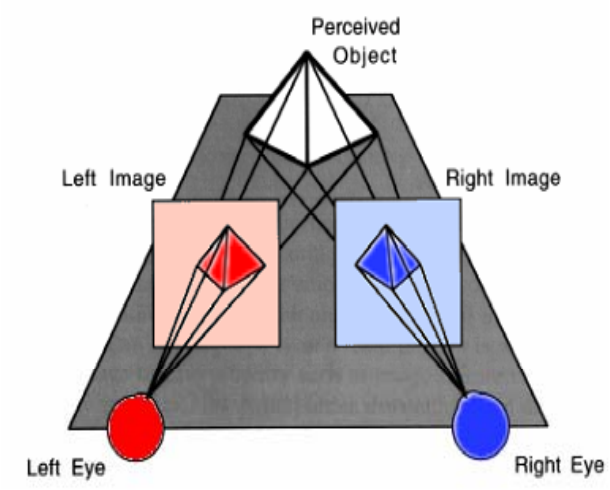

Figure 2.2.1. Human Vision System

The basic requirements for stereo vision are two 2D images which are captured from slightly different positions and calibrated cameras (inner and outer orientation).

The advantages of stereo vision technique includes:

- No need for active sources for energy emission as it uses the normal light in the scene and most importantly, light source position is not directly involved in the triangulation calculation.
- No need for moving parts as the objects are already covered in the images.
- Can be used in almost everywhere as no light conflict expected (as in active vision).
- Can be used for a wide range of object sizes and distances (even in satellite images).

The main problem in stereo vision is finding the correct corresponding points in the images used. This problem is still unsolved completely so, we still have limitations on practical applications of stereo vision mainly due to the matching problem.

2.3 Stereo vision applications

Obtaining 3D perception of the objects from 2D images allows us to measure distances, shapes, volumes and many other measures which are useful in practical world. Stereo vision has a wide range of application due to its advantages. Some of possible applications of stereo vision are:

- 3D inspection which can be used in many aspects such as quality control, deformation analysis,..etc.
- Constructing 3D models for virtual reality and online web applications.
- Automatic acquisition of 3D information to be used in CAD-CAM systems.
- Automatic creation of human models for medical, biomedical and bioengineering applications.

2.4 Stereo Vision process

The typical stereo vision process contains three main stages:

1. Preprocessing.
2. Matching.
3. Depth recovery.

Preprocessing

In the preprocessing stage, there are two main tasks. The first task is to identify image primitives that will be used in the matching stage. The primitives depend on which matching technique will be used. Using area-based technique, primitives can be described by intensity values of the image. On the other hand, features (like edges) can be used as primitives in a feature –based technique.

The second task is finding a point of interest to start the matching process. This selection is achieved by using interest operators. Again, these interest operators differes according to primitives and matching techniques. For example, in area-based technique, Moravec[2] suggests an interest operator depends on local maxima of directional variance measure over specific window around a point. For feature-based techniques, some interest operators are proposed like derivative operators, convolution, local gray-level intensity operators.

Matching

Matching stage concerns about finding the corresponding points between the two 2D images.

It is a critical stage in stereo vision computation as it seriously affect the next stage. Matching algorithms can be classified according to either matching primitives and/or imaging geometry.

According to matching primitives, we have:

- Area-based (also known as intensity-based) in which points or blocks of the image are used as primitives. Also the comparison is done directly on intensity values at these primitives.
- Feature-based in which some features (like edges) are used as primitives. Comparison is done on these features not on the intensity values.

According to imaging geometry, we have:

- Parallel-axis stereo where the cameras used to capture images are aligned such that their optical axis are parallel.
- Non-parallel (perspective) stereo where the optical axes of used cameras are not parallel.

- Number of cameras [binocular: two cameras; trinocular: three cameras; or multicular: more than three cameras]

In this project, we are concerning about the area-based technique. So, primitives used are blocks of the image.

Depth recovery

Observing human visual system, it has been found that the eyes provides two slightly different 2D images. Due to that difference, the same point is located in slightly different location in each image. The difference between the two locations of same real point is called the "disparity". Disparity is an important term in stereo vision as it represents the relative depth of the point (as shown in figure 2.4.1). Then if other parameters are known (such as camera focal length 'f' and distance between the two cameras used 'B') the real depth can be recovered.

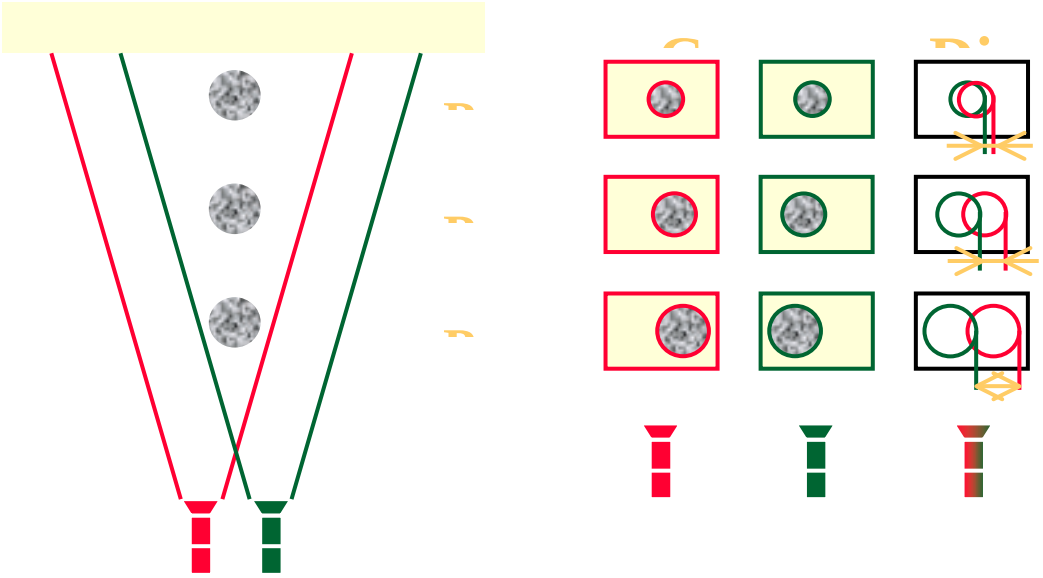

Figure 2.4.1 Disparity and its relation with relative depth

The other parameters needed to recover the real depths depend on the acquisition setup. Images can be captured with one, two, or more cameras. The parallel axis system with two cameras is the most common approach (as shown in figure 2.4.2). Alternatively, we can use only one camera to capture the images. In the one camera approach, first image is taken, camera is moved along only one axis (most propably the horizontal axis) for a small distance, then the second image is taken.

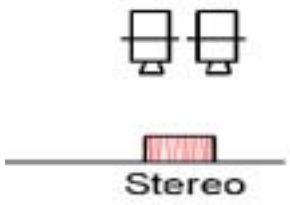

Figure 2.4.2 Parallel-axis system with two cameras

3. Methodology

Test images were used in the experiments that will be described below. The main target was to compute disparity maps. Since no calibration procedure happened the disparity maps represent relative disparity maps.

3.1 Test images with one corresponding point.

Test images were created by computing 15 x 15 size matrices. All the elements of the left image (A_{ij}) were set equal to zero apart from 5 randomly chosen pixels that inlcuded values from 1 until 5. The right image (B_{mn}) was created with the same way but the corresponding pixels to those in the right image were shifted by 4 pixels across the *j* direction. The test images used are shown in figure 3.1.1.

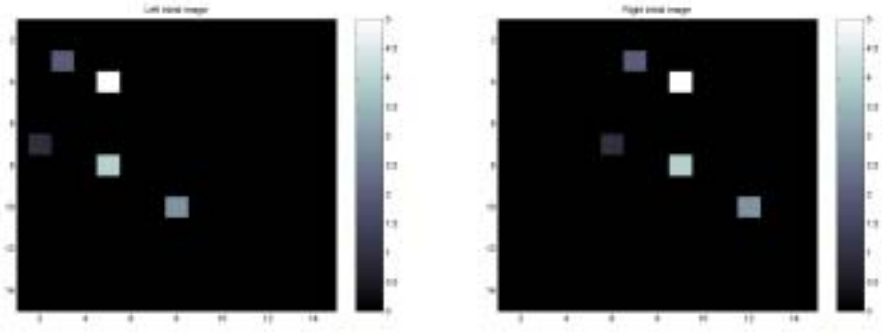

Figure 3.1.1 The left and the right test images. The black background represents all the elements with values equal to zero. The colorcode identify pixels with different values. The order in the right image is the same but shifted by 4 pixels.

As it can be seen from the two images in figure 3.1.1, each chosen point in the left image has only one corresponding point in the right image. The algorithm computed in order to calculate the disparity map in the case of one matching point is described below:

1. Define the window size $winA_{ij}$ in the left image A_{ij}.
2. Set a window, $winB_{mn}$, with equal area to *winAij* in the right image B_{mn} and scan through the whole image.
3. If the window size is equal to one pixel size, then
    a. If $winA_{ij}$ and $winB_{mn}$ are equal to zero, then set the correlation R_{mn} equal to zero and exit, otherwise
    b. calculate the correlation between $winA_{ij}$ and $winB_{mn}$ and map the correlation values to the right image.
    c. If correlation R_{mn} is equal to 1, then the position of $pixel_{mn}$ is stored.
    d. Compute the relative disparity, D_{ij} by calculating the absolute difference between the corresponding points ($D_{ij}=|j-n|$) and map the values to the left image.
4. If the window size is greater than one, then
    a. scan the right image B_{mn} with step equal to window size $winA_{ij}$.
    b. Calculate the maximum value of each window, $maxwinA_{ij}$ and $maxwinB_{mn}$.
    c. Repeat steps 3a-3d, but this time comparing $maxwinA_{ij}$ and $maxwinB_{mn}$.
5. Display the correlation in 2D and 3D plots and the relative disparity in 3D plot.

3.2 Test images with multiple candidate corresponding points.

In reality there are many matching points between the left and the right images. Therefore, it was necessary to compute an algorithm that is going to choose a corresponding point from the right image with the highest confident level. Test images of size 3 x 11 were created. All elements were set equal to 255 apart from 2 elements with values equal to zero. The right image was generated the same way and the two points with value equal to zero were shifted by two pixels. The two new test images are shown in figure 3.2.1.

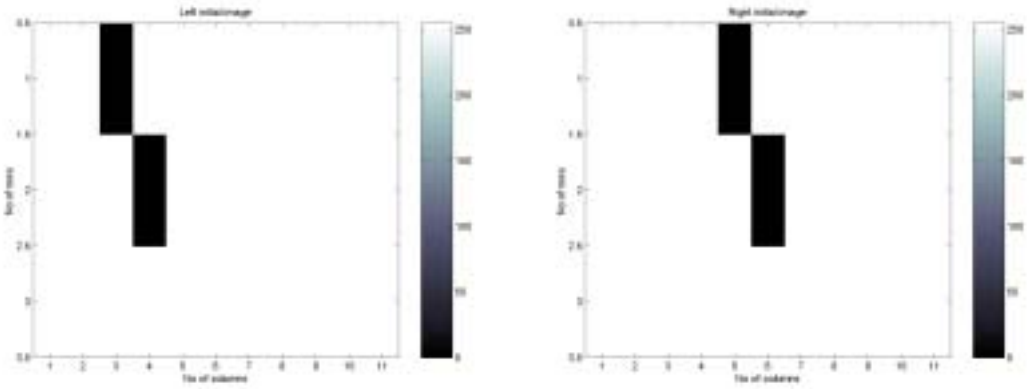

Figure 3.2.1 Test images with multiple corresponding points. The white background represent all the elements with values equal to 255 and the black areas are the pixels with values 0.

The algorithm described in section 3.1 needed some modifications in order to recognise which point of the right image corresponds to the left image, since for example for the first black pixel, it will find two answers – two corresponding points. Therefore, it was necessary to chose a procedure of selecting the corresponding points with high confidence level.

The modifications in the algorithm mentioned in section 3.1 are:

1. The scanning process in the right image occurs not through the whole image but only through the same row.
2. The correlation is computed following the same steps as in the initial algorithm.
3. Then the nearest point with the maximum correlation value is chosen and its position is stored.
4. The disparity is calculated as before.

4. Results

4.1 Test images with one corresponding point.

Different window sizes were used in order to compute the relative disparity map. Starting from the simplest one, which is one pixel size, the correlation was one between the corresponding (matcing) points. Figure 4.1.1 shows the plots of the correlation in 2 and 3 dimensions.

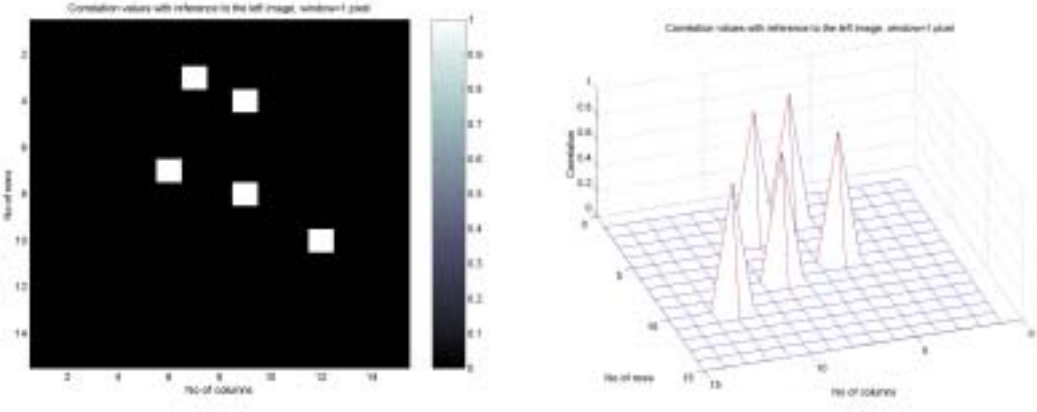

Figure 4.1.1. The calculated correlation function computed between the left and right images with reference to the left with window size equal to one pixel. Where the corresponding points are the correlation is one, otherwise is zero. The 3D plot was rotated so it can match the 2D plot for better visualisation.

The relative disparity map between the corresponding points is calculated and presented in figure 4.1.2. Since the shift between the two images is the same, it was expected that the disparity values are the same for all points.

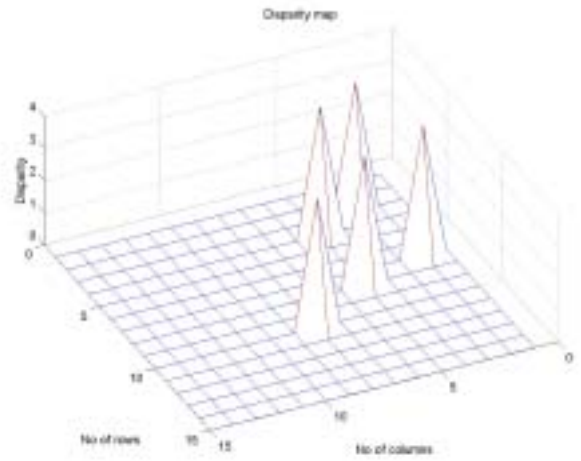

Figure 4.1.2 The relative disparity map using window size equal to one pixel. The values of the calculated disparity were mapped to the left image.

It can be observed from figure 4.1.2 that the disparity values are the same, as it was expected because of the same shift by 4 pixels. The peaks are shown in the right upper corner because the axes in the 3D plot are reversed. However, the peaks do correspond to the positions of the chosen points ine the left image.

The window size changed from one pixel to 2 x 2 pixels. The algorithm is the one mentioned in section 3.1. The maximum correlation value is mapped to the pixel of the upper left corner of the window and so did the calculated relative disparity. The results of the correlation in 2D and 3D plots are displayed in figure 4.1.3 and the relative disparity map in figure 4.1.4.

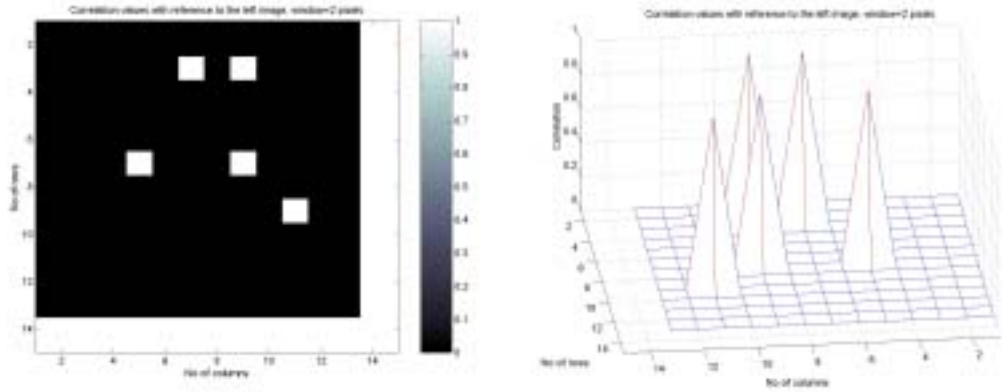

Figure 4.1.3 The calculated correlation function computed between the left and right images with reference to the left with window size 2x2 pixels. Where the corresponding points are the correlation is one, otherwise is zero. The 3D plot was rotated so it can match the 2D plot for better visualisation.

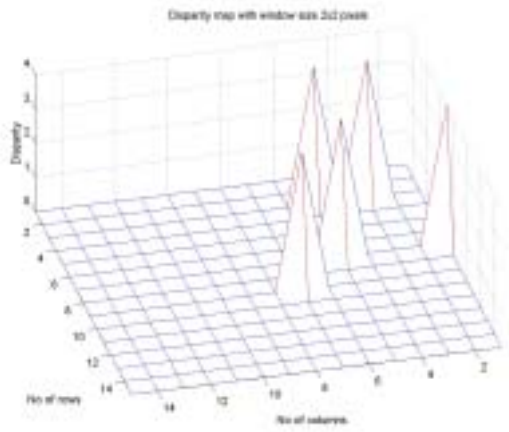

Figure 4.1.4 The relative disparity map using window size 2x2 pixels. The values of the calculated disparity were mapped to the left image.

The white borders in figure 4.1.3 are due to limitations because of the window dimensions. The relative disparity values were mapped inside the window, which means that in some cases the correlation was found one in the adjacent pixel of the corresponding point. This was an expected result commonly appeared in cases that the window has dimensions greater than one pixel.

4.2 Test images with multiple candidate corresponding points.

The algorithm for the multiple corresponding points was checked in the images shown in figure 3.2.1. The window size is equal to one. The results for the correlation and the disparity are show in figures 4.2.1 and 4.2.2 respectively.

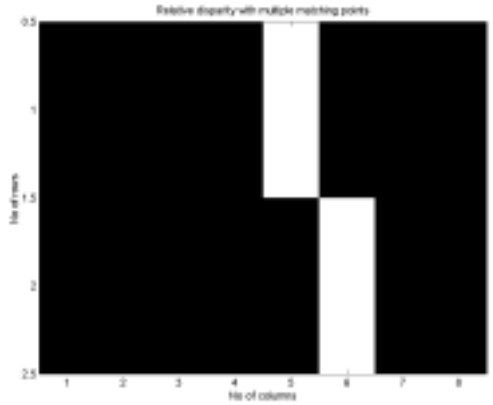

Figure 4.2.1 The correlation function in the presence of multiple corresponding points. The black background represents correlation zero and the white pixels correlation one.

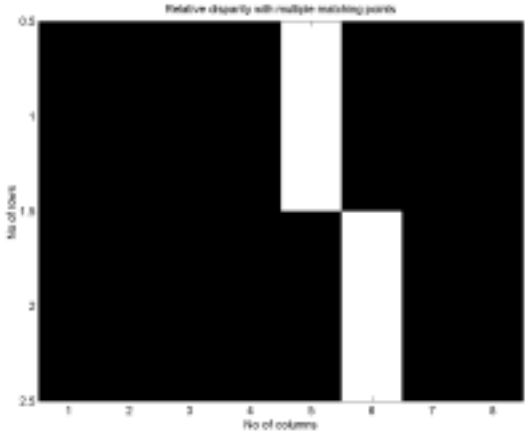

Figure 4.2.2 The disparity 2D map in the presence of multiple corresponding points. The window size is equal to one pixel size.

It can be observed that the relative disparity was calculated only in the locations that the respective corresponding points were found. Since the window size was equal to the pixel size there was no uncertainty in defining the exact position of the corresponding point.

5. Discussion - Conclusions

The test images tried here represent only simple - optimal situations. In reality the task of computing disparity maps between two stereo images is much more difficult. During the first set of experiments, when the chosen point in the left image had only one corresponding point in the right image, it was an introductory experiment in order to understand and practice the computations of how to locate the point in the left image, how to define exactly the same area in the right image, how to compute the correlation and finally how to calculate the relative disparity between the corresponding points. When the window size changed to 2x2 pixels, it was found that the uncertainty in estimating the corresponding point position is equal to the dimension of the window (referring to square windows). This is the reason why two of the

corresponding points were found within the window dimension of the exact position. The experiment was repeated for greater window sizes and the above observation was verified.

One of the difficulties in stereo images is the fact that many corresponding points might be found between one point of the left and the right image. Therefore, there must be a selection criterion that chooses the corresponding point with the higher confidence level. In the second set of experiments presented here, this situation was examined with two candidate corresponding points in the right image for each point of the left image. The selection algorithm was based in the fact that the candidate with higher confidence is the nearest point in the same row as the point if th eleft image. This algorithm performed very well in the case described in section 3.2 as it was verified from the results presented in section 4.2.

Figure 5.1 The two stereo images of landscape. The team tried to work with those two but many difficulties appeared.

Concluding this work, we must report that there were many difficulties that the team had to face when we tried to work with real stereo images from landscape or people (figure 5.1). The images were initially very big (1600 x 1600) and the calculations were very slow. Additionally, since no member had previous experience it was difficult to spot the key-points when all the attention should have been focused. Therefore, the solution of simple test images gave us the opportunity to understand more the process of calculating disparity maps. Because of time restrictions, the work had to terminate by showing the results of the test images and tasting a little bite from the world of stereo images.

6. Bibliography

1. Barnard S T, Thomson W B, “Disparity Analysis of Images”, IEEE Tran. Pattern Analysis and Machine Intelligence, Vol 2, No 4, pages: 333 – 340, 1980.

2. Dhond U R, Aggarwal J K, “Structure from Stereo – A Review”, IEEE Tran. Systems, Man and Cybernetics, Vol.19, No 6, pages: 1489 – 1510, 1989.

3. Klette R, Karsten Schluns, and Andreas Koschan, “Computer Vision: Three-Dimensional Data from images”, Springer-Verlag Singapore Pte. Ltd., 1998.

4. Marr D, “Vision”, Freeman and Company, 1982.

5. Prazdny K, ”Detection of Binocular Disparities”, Biological Cybernetics, Vol. 52, pages 93-99, 1985.

6. Suresh B. Marapane and Mohn M. Trivedi, ”Region-Based Stereo Analysis for Robotic Applications”, IEEE Transactions on Systems, Man, and Cybernetics, Vol. 19, No. 6, pages 1447-1464, 1989.